National Library
of Canada

Canadian Theses Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale
du Canada

Services des thèses canadiennes

CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NL-339 (r.86/06)

Canada

THE UNIVERSITY OF ALBERTA

C-TESTS, PROFICIENCY, AND READING STRATEGIES IN ESL

by

© SOMSAK BOONSATHORN

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELEMENTARY EDUCATION

EDMONTON, ALBERTA
SPRING, 1987

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-37731-3

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR SOMSAK BOONSATHORN

TITLE OF THESIS C-TESTS, PROFICIENCY, AND READING STRATEGIES IN ESL

DEGREE FOR WHICH THESIS WAS PRESENTED DOCTOR OF PHILOSOPHY

YEAR THIS DEGREE GRANTED SPRING, 1987

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(SIGNED) Somsak Boonsath

PERMANENT ADDRESS:

Nakhonsawan Teachers' College
Nakhon Sawan 60000
Thailand

DATED April 13, 1987

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled C-TESTS, PROFICIENCY, AND READING STRATEGIES IN ESL submitted by SOMSAK BOONSATHORN in partial fulfilment of the requirements for the degree of DOCTOR OF PHILOSOPHY in TEACHING ENGLISH AS A SECOND LANGUAGE.

Supervisor

Douglas V. Parker

R. K. Jackson

Grace Malicky

Terry Piper

External Examiner

Date 13 April 1987

TO M_ MOTHER, M_ FATHER, A__

ALL __O HAVE ____ATED AND _____RAGED ME.*

---

*The first line is in the form of the C-Test and the second the MC-Test. (To my mother, my father, and all who have educated and encouraged me.)

ABSTRACT

This study attempted to verify the reliability and validity of the C-Test which has been used to resolve issues confronting the conventional cloze test. In the C-Test, the respondent is asked to restore reading passages in which the second half of every second word is deleted. A modified version of the C-Test, namely the MC-Test where the first half of the words is deleted, was developed to compare with the C-Test.

Two forms of the C-Test and two forms of the MC-Test were administered to a total of 389 native English speakers and 104 ESL learners. The ESL subjects also wrote the Michigan Test which was used as a criterion measure, and 28 of them were chosen for oral interviews.

Test statistics (item difficulty, item discrimination, and test reliability) were computed from both L1 and L2 data. The L1 data were factor analyzed and correlations between the C-Test/MC-Test and the Michigan Test were calculated from the L2 data. In addition, an error analysis was performed on the C-Test/MC-Test responses of the 28 ESL interviewees.

The results of the investigations indicated that the newly-proposed format, the MC-Test, was superior to the C-Test in several respects. The MC-Test discriminated better, had a greater reliability, and was more valid with regard to factor structures, than the C-Test. In reference to factor structures, it could be argued that the MC-Test requires more reading strategies than does the C-Test. While for the C-Test, respondents appeared to rely primarily on the orthographic cues available to them, the restoration of the MC-Test passages seemed to necessitate the use of all the strategies which readers employ in the process of normal reading. Error analysis indicated that proportionately the categorized errors identified in the C-Test and the MC-Test were different. The interviews revealed that the high proficiency group and the low proficiency group used similar strategies in processing the C-Test/MC-Test passages in spite of their different reader types.

ACKNOWLEDGEMENTS

I wish to express my gratitude to the numerous individuals who have made the completion of this study possible.

I am especially indebted to Dr. Warren D. Wilde, my dissertation supervisor, for his guidance, encouragement, understanding, and continued support. His supervisory expertise and talent were invaluabe. Appreciation is sincerely extended to Dr. Thomas O. Maguire and Dr. Robert K. Jackson, my supervisory committee members, who provided insightful statistical and organizational advice and assistance throughout all phases of the study.

I am also grateful to Dr. Grace V. Malicky, Dr. Douglas V. Parker, the members of the examining committee, and Dr. Terry Piper, the external examiner from the University of Calgary. Their contributions to the improvement of the study are fully acknowledged. In addition, my thanks are given to Dr. W. T. Fagan, Dr. P. A. McFetridge, and Dr. J. T. Patrie, for their initial encouragement and very useful advice.

In particular, I would like to thank the staff and the students of the co-operative institutions for their assistance and willingness to participate. I am grateful to the Departments of Elementary Education and Linguistics at the University of Alberta which provided financial support all through my study program, and to the Government of Thailand which granted me study leave.

I also wish to thank my friends, A. E. Mackie, M. L. Smith, and others whose names are not listed, for their friendship, assistance, and moral support.

A special note of thanks is extended to my family, Arporn, Wasita, and Sittha, for their unique kind of encouragement.

Table of Contents

Chapter Page

1. INTRODUCTION AND BACKGROUND 1
  - 1.1 CLOZE PROCEDURE 1
  - 1.2 THE PROBLEM 4
    - 1.2.1 DEVELOPMENT OF THE C-TEST 4
    - 1.2.2 THE MODIFIED C-TEST 5
  - 1.3 STATEMENT OF THE PROBLEM 7
  - 1.4 PURPOSE OF THE STUDY 8
  - 1.5 RESEARCH QUESTIONS 8
  - 1.6 THEORETICAL RATIONALE 9
  - 1.7 ASSUMPTIONS UNDERLYING THE C-TEST 12
  - 1.8 DEFINITION OF TERMS 13
2. REVIEW OF RELATED LITERATURE 15
  - 2.1 RESEARCH ON CLOZE PROCEDURE 15
    - 2.1.1 METHODOLOGICAL CONSIDERATIONS 15
      - 2.1.1.1 FREQUENCY OF DELETIONS 15
      - 2.1.1.2 TYPE OF DELETIONS 19
      - 2.1.1.3 NUMBER OF DELETIONS 20
      - 2.1.1.4 SCORING METHODS 21
    - 2.1.2 CLOZE PROCEDURE AS A MEASURE OF ESL/EFL PROFICIENCY 23
    - 2.1.3 THE MODIFIED CLOZE PROCEDURE 26
      - 2.1.3.1 MULTIPLE-CHOICE CLOZE TESTS 26
      - 2.1.3.2 MATCHING CLOZE TESTS 28
      - 2.1.3.3 WELL-TAILORED CLOZE TEST 29
      - 2.1.3.4 C-TESTS 30
  - 2.2 RESEARCH ON READING STRATEGIES 32
    - 2.2.1 THE READING PROCESS 32

2.2.2 READING STRATEGIES 33
2.2.3 STUDIES IN SECOND LANGUAGE READING 34
3. PROCEDURE 39
3.1 TEST CONSTRUCTION 39
3.1.1 SELECTION OF PASSAGES 39
3.1.2 MUTILATION 39
3.1.3 SCORING METHOD 40
3.2 CRITERION MEASURES 40
3.3 SAMPLES OF THE STUDY 41
3.4 DATA COLLECTION 41
3.5 DATA ANALYSIS 43
3.6 PILOT STUDY 44
3.6.1 PART I 44
3.6.2 PART II 44
4. FINDINGS AND INTERPRETATION 49
4.1 COMPARABILITY 49
4.1.1 ITEM DIFFICULTY 49
4.1.1.1 OVERALL RESULTS (L1 DATA) 49
4.1.1.2 SUBTEST RESULTS (L1 DATA) 53
4.1.1.3 OVERALL RESULTS (L2 DATA) 53
4.1.1.4 SUBTEST RESULTS (L2 DATA) 58
4.1.2 SUMMARY OF ITEM DIFFICULTY 60
4.1.2.1 STRUCTURE OF FUNCTION WORDS 60
4.1.2.2 ANOMALOUS ITEMS 63
4.1.3 ITEM DISCRIMINATION 66
4.1.3.1 L1 DATA 66
4.1.3.2 L2 DATA 68

4.1.4 SUMMARY OF ITEM DISCRIMINATION 70
4.2 RELIABILITY 71
4.2.1 L1 DATA 71
4.2.2 L2 DATA 73
4.2.3 SUMMARY OF RELIABILITY 75
4.3 VALIDITY 76
4.3.1 FACTOR ANALYSIS (L1 DATA) 76
4.3.1.1 ADJACENCY EFFECT 77
4.3.1.2 CONTENT WORDS VS. FUNCTION WORDS 79
4.3.2 FACTOR SOLUTIONS 81
4.3.2.1 C-TEST 1 SUBTEST 1 (C1S1) 82
4.3.2.2 MC-TEST 1 SUBTEST 1 (MC1S1) 86
4.3.2.3 COMPARISON OF FACTOR STRUCTURES (C1S1 & MC1S1) .. 87
4.3.2.4 C-TEST 1 SUBTEST 2 (C1S2) 89
4.3.2.5 MC-TEST 1 SUBTEST 2 (MC1S2) 93
4.3.2.6 COMPARISON OF FACTOR STRUCTURES (C1S2 & MC1S2) .. 93
4.3.2.7 C-TEST 1 SUBTEST 3 (C1S3) 95
4.3.2.8 MC-TEST 1 SUBTEST 3 (MC1S3) 95
4.3.2.9 COMPARISON OF FACTOR STRUCTURES (C1S3 & MC1S3) .. 99
4.3.2.10 C-TEST 1 SUBTEST 4 (C1S4) 99
4.3.2.11 MC-TEST 1 SUBTEST 4 (MC1S4) 104
4.3.2.12 COMPARISON OF FACTOR STRUCTURES (C1S4 & MC1S4) 105
4.3.2.13 C-TEST 2 SUBTEST 1 (C2S1) 105
4.3.2.14 MC-TEST 2 SUBTEST 1 (MC2S1) 110
4.3.2.15 COMPARISON OF FACTOR STRUCTURES (C2S1 & MC2S1) 111
4.3.2.16 C-TEST 2 SUBTEST 2 (C2S2) 111
4.3.2.17 MC-TEST 2 SUBTEST 2 (MC2S2) 116

4.3.2.18 COMPARISON OF FACTOR STRUCTURES (C2S2 & MC2S2) 117
4.3.2.19 C-TEST 2 SUBTEST 3 (C2S3) 117
4.3.2.20 MC-TEST 2 SUBTEST 3 (MC2S3) 122
4.3.2.21 COMPARISON OF FACTOR STRUCTURES (C2S3 & MC2S3) 122
4.3.2.22 C-TEST 2 SUBTEST 4 (C2S4) 124
4.3.2.23 MC-TEST 2 SUBTEST 4 (MC2S4) 124
4.3.2.24 COMPARISON OF FACTOR STRUCTURES (C2S4 & MC2S4) 128
4.3.3 SUMMARY OF FACTOR ANALYSIS 128
4.3.4 CORRELATIONAL ANALYSIS (L2 DATA) 132
4.3.5 SUMMARY OF CORRELATIONAL ANALYSIS 134
4.4 ERROR ANALYSIS AND INTERVIEWS 134
4.4.1 ERROR ANALYSIS 135
4.4.1.1 C-TEST ERROR INVENTORY 135
4.4.1.2 OVERALL RESULTS 136
4.4.1.3 SUBSTITUTION ERRORS 138
4.4.2 SUMMARY OF ERROR ANALYSIS 148
4.4.3 INTERVIEWS 151
4.4.3.1 ATTITUDE TOWARD THE TEST 151
4.4.3.2 FACE VALIDITY 153
4.4.3.3 READING STRATEGIES 153
4.4.4 SUMMARY OF INTERVIEWS 155
5. SUMMARY, CONCLUSIONS, AND IMPLICATIONS 156
5.1 SUMMARY OF FINDINGS 156
5.1.1 COMPARABILITY 156
5.1.2 RELIABILITY 157
5.1.3 VALIDITY 157
5.1.4 ERROR ANALYSIS AND INTERVIEWS 157

5.2 CONCLUSIONS AND DISCUSSION 158
5.3 IMPLICATIONS 162
5.3.1 LIMITATIONS OF THE STUDY 162
5.3.2 IMPLICATIONS FOR EDUCATION 163
5.3.3 SUGGESTIONS FOR FURTHER RESEARCH 164
BIBLIOGRAPHY 165
APPENDICES 173

APPENDIX A

TESTS AND GUIDELINES FOR ADMINISTRATION 174
GUIDELINES FOR TEST ADMINISTRATION 175
C-TEST 1 178
MC-TEST 1 185
C-TEST 2 192
MC-TEST 2 199

APPENDIX B

ITEM STATISTICS 206

APPENDIX C

CROSS-TABULATIONS OF ITEM DISCRIMINATION 215

APPENDIX D

FACTOR LOADING MATRICES 224

APPENDIX E

C-TEST ERROR CODING SHEET 233

APPENDIX F

SAMPLE INTERVIEW QUESTIONS 236

List of Figures

Figure Page
4.1 Equipercentile Plot of C-Test 1 and MC-Test 1: L1 Data 51
4.2 Equipercentile Plot of C-Test 2 and MC-Test 2: L1 Data 52
4.3 Equipercentile Plot of C-Test 1 and MC-Test 1: L2 Data 56
4.4 Equipercentile Plot of C-Test 2 and MC-Test 2: L2 Data 57
4.5 C1S1 84
4.6 MC1S1 85
4.7 C1S2 91
4.8 MC1S2 92
4.9 C1S3 97
4.10 MC1S3 98
4.11 C1S4 102
4.12 MC1S4 103
4.13 C2S1 108
4.14 MC2S1 109
4.15 C2S2 114
4.16 MC2S2 115
4.17 C2S3 120
4.18 MC2S3 121
4.19 C2S4 126
4.20 MC2S4 127

List of Tables

Table Page

3.1 Distribution of Grade Levels and Test Forms for L1 Sample 42

3.2 Distribution of Language Backgrounds and Test Forms for L2 Sample 42

3.3 Mean Percent Scores and S.D. of Tests (L1 Data: Pilot Study) 45

3.4 Mean Percent Scores and S.D. of Tests (L2 Data: Pilot Study) 47

3.5 Comparison between the C-Test/MC-Test Scores and the Michigan Test Scores of L2 Subjects 47

4.1 Mean Percent Scores and S.D. of the C-Tests and MC-Tests (L1 Data) 50

4.2 Mean Percent Scores and S.D. of C-Test and MC-Test Subtests (L1 Data) 54

4.3 Mean Percent Scores and S.D. of the C-Tests and MC-Tests (L2 Data) 55

4.4 Mean Percent Scores and S.D. of the C-Test and MC-Test Subtests (L2 Data) 59

4.5 Orthographic Structures of Function Words 62

4.6 Items for Which the Difficulty Indices in C-Test 1 are Smaller than in MC-Test 1 64

4.7 Items for Which the Difficulty Indices in C-Test 2 are Smaller than in MC-Test 2 65

4.8 Summary of Item Discrimination Statistics for C-Test 1 and MC-Test 1 (L1 Data) 67

4.9 Summary of Item Discrimination Statistics for C-Test 2 and MC-Test 2 (L1 Data) 67

4.10 Summary of Item Discrimination Statistics of C-Test 1 and MC-Test 1 (L2 Data) 69

4.11 Summary of Item Discrimination Statistics of C-Test 2 and MC-Test 2 (L2 Data) 69

4.12 Summary of Reliability Statistics of C-Tests and MC-Tests (L1 Data) 72

4.13 Summary of Reliability Statistics of C-Test and MC-Test Subtests (L1 Data) 72

4.14 Summary of Reliability Statistics of C-Tests and MC-Tests (L2 Data) 74

Table Page

4.15 Summary of Reliability Statistics of C-Test and MC-Test Subtests (L2 Data) 74

4.16 Number of Factors Extracted from Each Subtest of the C-Tests and MC-Tests 78

4.17 The Major Word Classes in English 80

4.18 Items Loading 0.4 or Greater on Each Factor: Subtest 1 (C1S1 vs. MC1S1) 83

4.19 Cross-tabulation of Items Loading on Each Factor of C1S1 and MC1S1 88

4.20 Items Loading 0.4 or Greater on Each Factor: Subtest 2 (C1S2 vs. MC1S2) 90

4.21 Cross-tabulation of Items Loading on Each Factor of C1S2 and MC1S2 94

4.22 Items Loading 0.4 or Greater on Each Factor: Subtest 3 (C1S3 vs. MC1S3) 96

4.23 Cross-tabulation of Items Loading on Each Factor of C1S3 and MC1S3 100

4.24 Items Loading 0.4 or Greater on Each Factor: Subtest 4 (C1S4 vs. MC1S4) 101

4.25 Cross-tabulation of Items Loading on Each Factor of C1S4 and MC1S4 106

4.26 Items Loading 0.4 or Greater on Each Factor: Subtest 1 (C2S1 vs. MC2S1) 107

4.27 Cross-tabulation of Items Loading on Each Factor of C2S1 and MC2S1 112

4.28 Items Loading 0.4 or Greater on Each Factor: Subtest 2 (C2S2 vs. MC2S2) 113

4.29 Cross-tabulation of Items Loading on Each Factor of C2S2 and MC2S2 118

4.30 Items Loading 0.4 or Greater on Each Factor: Subtest 3 (C2S3 vs. MC2S3) 119

4.31 Cross-tabulation of Items Loading on Each Factor of C2S3 and MC2S3 123

4.32 Items Loading 0.4 or Greater on Each Factor: Subtest 4 (C2S4 vs. MC2S4) 125

Table Page

4.33 Cross-tabulation of Items Loading on Each Factor of C2S4 and MC2S4 129

4.34 Correlations between the Michigan Test and the C-Test/MC-Test 133

4.35 Errors Identified in High and Low Proficiency Groups 137

4.36 Errors Identified in C-Test and MC-Test Formats 137

4.37 Substitution Errors Identified in High and Low Proficiency Groups 139

4.38 Substitution Errors Identified in C-Test and MC-Test Formats 139

4.39 Comparison of Errors for C1S1 and MC1S1 140

4.40 Comparison of Errors for C1S2 and MC1S2 141

4.41 Comparison of Errors for C1S3 and MC1S3 142

4.42 Comparison of Errors for C1S4 and MC1S4 143

4.43 Comparison of Errors for C2S1 and MC2S1 144

4.44 Comparison of Errors for C2S2 and MC2S2 145

4.45 Comparison of Errors for C2S3 and MC2S3 146

4.46 Comparison of Errors for C2S4 and MC2S4 147

4.47 Summary of Interview Data 152

Chapter 1

INTRODUCTION AND BACKGROUND

1.1 CLOZE PROCEDURE

A tremendous amount of research has been performed on cloze procedure since it was pioneered by Taylor (1953), who proposed it as "a new psychological tool for measuring the effectiveness of communication." As Taylor (1953) himself remarked, the procedure was initially investigated as a new approach for estimating readability. However, he foresaw that "this tool seems likely to have a variety of applications, both theoretical and practical, in other fields involving communication functions" (p. 415).

Cloze testing has been inspired from Gestalt psychology which hypothesizes that the mind can predict the totality of an incomplete form or shape. A familiar example of this concept is one's ability to see a broken circle as a whole one by mentally closing up the gaps. The ability to make such completion in one's mind is called "closure" by Gestalt psychologists who contend that the process of learning consists of global comprehension first to be followed later by the comprehension of detail (Stansfield & Hansen, 1983, p, 30).

The cloze technique of test construction is very simple. After selecting a passage for the purpose, the test constructor simply deletes every *n*th word (e.g., 5th, 6th, or 7th), normally starting from the second sentence and leaving the last sentence unmutilated. The student is then required to replace the missing words or restore the passage by placing other contextually acceptable words in the blanks.

Below, an example of a cloze test is included, taken from the researcher's (Boonsathorn, 1977, p. 93) master's thesis with a slight modification.

CLOZE TEST

DIRECTIONS: In the following passage, there are 30 blanks for you to fill in. You should:

1. read the whole passage through to get the general idea;
2. go back and fill in the blank with the words you think are missing;
3. use only one word for each blank; and
4. try to fill in every blank in 20 minutes.

There was once an officer who had plenty of money and always bought beautiful clothes, but he was prouder of his boots than of anything else that he wore.

This officer had a very good _____, who always kept his clothes very _____ and tidy and made his boots _____ more brightly than those of any _____ the other officers, but this servant _____ old, and one day he had _____ retire and let another soldier take _____ place.

The officer's new servant was _____ and clever, but lazy. One morning, _____ it had rained a lot during _____ night before, and the roads were _____ muddy, the officer saw that his _____ had not been cleaned, so he _____ his new servant and said to _____ , "I am going to go for _____ ride this morning, and my boots _____ not been cleaned."

"Sir," answered the _____ servant very politely, "it rained a _____ last night, and there is a _____ of mud on the roads this _____."

"Yes," answered the officer, "I agree, _____ what has that got to do _____ my boots?"

"Well, sir," explained the _____, "if I clean your boots now, _____ will soon get dirty again, so _____ is a waste of time to _____ them."

The officer said nothing, but _____ he had returned from his ride _____ had eaten his lunch, he did _____ leave any food for his servant _____ he usually did.

When the servant asked why he was not given any food for lunch, the officer calmly said, "It would be a waste of time to give you any food now, because you would only be hungry again in a few hours' time."

The research on cloze procedure during the first few years after its initiation in 1953 appeared to be limited to only the studies and writings by Taylor and a few other researchers, namely, Osgood, Wilson, and Carroll. After that short period of time, however, the development of research on cloze procedure gained considerable momentum (Anderson, 1976, p. 21). Cloze procedure was first applied to measure the level of difficulty of passages for reading in the first language. Numerous studies were completed on this application. Subsequently, it was used to test first language reading comprehension. Finally, the cloze procedure was adapted for foreign language testing as a device for evaluating global competence in the target language. Oller (1973b), one of the outstanding researchers in the area, claims that: "One of the most promising types of integrative skills tests which has been proposed for measuring either achievement or proficiency in foreign language or second language situations is the cloze test" (p. 192). Many of the studies on this aspect have focused on one or both of the main characteristics of a test: reliability and validity (Brown, 1984, p. 109).

The use of cloze procedure in testing appears to conform to Goodman's (1967) model of reading process. Goodman proposes that:

> More simply stated, reading is a psycholinguistic guessing game. It involves an interaction between thought and language. Efficient reading does not result from precise perception and identification of all elements, but from skill in selecting the fewest, most productive cues necessary to produce guesses which are right the first time. The ability to anticipate that which has not been seen, of course, is vital in reading, just as the ability to anticipate what has not yet been heard is vital in listening. (reprinted in Gollasch (Ed.), 1982, pp. 33-34)

Goodman contends that while reading aloud, readers encode the print for meaning and recode it for oral language. They are able to follow the sequence to anticipate the outcome through orthography, syntax, and semantics. Redundancies in these three areas further assist good readers in confirming their predictions (Cooper & Petrosky, 1976, p. 193). Consequently, reading comprehension is enhanced by one's ability to interpret as many graphophonic, syntactic, and semantic cues as possible. As observed by Smith, Goodman, & Meredith (1970), "comprehension depends on the reader's using all the cues available to him" (p. 266).

Propst & Baldauf (1981) used evidence provided by Neville & Pugh (1976-77) to support Goodman's hypothesis about text sampling in the reading process and to support the rationale that the restoration of cloze passages is directly related to the reading process. In their research, it was evidenced that "poor readers use only the information in front of a blank in determining what word to write in particular blanks on a cloze test, while better readers use information before and after the blanks, thus enabling the better readers to more often determine the correct words to be inserted in the blanks on cloze tests" (Propst & Baldauf, 1981, p. 87).

1.2 THE PROBLEM

Enthusiastic about the favorable results of the cloze procedure, the first researchers were readily prepared to accept the tests as reliable and valid. In recent years, however, a number of studies (Alderson, 1979, 1983; Klein-Braley, 1981, 1983; Foley, 1983; Brown, 1983, 1984; and Klein-Braley & Raatz, 1984) have raised criticisms about the procedure. Klein-Braley and Raatz (1984) summarize the major criticisms found as follows:

> 1. systematic *n*th word deletion does not necessarily produce a random sample of the elements of the text;
> 2. different deletion rates and starting points applied to the *same* text produce tests which can differ very considerably in difficulty, reliability and validity;
> 3. particularly for homogeneous samples (classroom groups or monolingual groups) cloze tests tend to have unsatisfactory reliability and validity coefficients;
> 4. there are major problems with scoring since scorers are very much less than unanimous about the acceptability of individual solutions offered in acceptable scoring procedures, while the use of exact scoring produces extremely difficult and therefore very frustrating tests;
> 5. in factorial studies very often a separate cloze factor appears, and in Klein-Braley's convergent/discriminant validation study not even convergent validity (correlation between two cloze tests) could be demonstrated. (p. 135)

In addition, Klein-Braley & Raatz (1984) note the following four difficult technical problems:

> 1. the question of text selection in terms of suitability and difficulty for the sample envisaged;
> 2. the fact that examinees are only presented with one text which can mean that text content is a source of bias in the scores;
> 3. the observation that in L1 testing, examinees, even when they are adult educated native speakers, rarely obtain a perfect score (cf., Pugh and Blenkhorn, 1984) although at the same time adult educated native speakers involved in L2 testing are expected to be able to make accurate judgements of acceptability when they act as scorers;
> 4. the use of KR-20 for the determining of test reliability is probably impermissible since the test items (deletions) may not be independent of each other. (p. 135)

These four technical difficulties make the justification of the conventional cloze procedure problematic. As Klein-Braley & Raatz (1984) have suggested, further study on the procedure is required to develop an alternative which ensures a format compatible with the theory articulated by cloze proponents.

1.2.1 DEVELOPMENT OF THE C-TEST

The C-test was developed in Duisburg, Germany by Raatz & Klein-Braley in 1981 as an alternative for the conventional cloze test. Based on the questions and problems found in the standard cloze test, a list of six criteria for a new test was proposed:

> 1. it should use several different texts;
> 2. it should have at least 100 deletions;
> 3. adult native speakers should obtain virtually perfect scores;
> 4. the deletions should affect a representative sample of the text;
> 5. only exact scoring should be possible;
> 6. the test should have high reliability and validity. (Klein-Braley & Raatz, 1984, p. 136)

The newly developed test, the C-Test, is an attempt to find an alternative test that would meet these criteria. (It is normally comprised of four to six short passages constructed according to Klein-Braley & Raatz's *rule of 2*.) The deletion in each passage begins in the second sentence and the second half of every second word is deleted until the required number of mutilations is reached, while the text then continues to the end of the paragraph. The following is an example from Klein-Braley & Raatz (1984, p. 136).

THE C-TEST

> There are usually five men in the crew of a fire engine. One o_ them dri___ the eng___. The lea___ sits bes___ the dri___. The ot___ firemen s__ inside t__ cab o_ the fi__ engine. T__ leader h__ usually be__ in t__ Fire Serv____ for ma__ years. H_ will kn__ how t_ fight diff_____ sorts o_ fires. S_, when t__ firemen arr___ at a fire, it is always the leader who decides how to fight a fire. He tells each fireman what to do.

1.2.2 THE MODIFIED C-TEST

According to Goodman (1967), in the process of reading, the reader simultaneously engages in all levels of processing--graphophonic, syntactic, and semantic. Word identification and comprehension are instantaneous for some readers. Smith (1978) argues that comprehension precedes perception; however, no process can operate in total isolation. In general, as Anderson (1976) points out:

> From a psycholinguistic viewpoint, the reading comprehension process begins with the written language encoded by the writer. The reader views the graphic symbols as sequences of visual stimuli. The message is decoded or interpreted when the reader, utilizing all his relevant past experiences, which include his previous learning and his language development, takes advantage of the semantic and syntactic cues present in the written language. (p.16)

To decode the message, cloze procedure requires the reader to predict and reconstruct the mutilated language patterns by making the most likely replacements in the light of his language system and the cues that are available. Based on this psycholinguistic viewpoint, since it is difficult to investigate the reader's past experiences or world knowledge, the focus of this investigation into cloze procedure will be only on the three main components--graphophonic/orthographic, syntactic/structural, and semantic cues--as present in graphic symbols in the mutilated words.

Since, in the conventional cloze test, only a blank and no orthographic cues are present, the prediction the reader makes to reconstruct the mutilated language patterns will be based solely on the syntactic/structural and semantic cues available from the context before and after the blank (besides his/her past experiences or world knowledge). In the C-Test, where the second half of every second word is deleted, some orthographic cues are given. In this case, the reader will make full use of these orthographic cues in addition to the syntactic/structural and semantic ones.

The modified C-Test (The MC-Test) to be proposed in this study basically conforms to Klein-Braley & Raatz's *rule of 2*. The deletion starts in the second sentence, but, instead of the second half, the first half of every second word is deleted until the required number of mutilations is reached. The text then continues, as in the original C-Test, to the end of the paragraph. The following example of the MC-Test was constructed from the same passage used as an example for the C-Test.

THE MC-TEST

There are usually five men in the crew of a fire engine. One _f them ___ves the ___ine. The ___der sits ___ide the ___ver. The ___er firemen _t inside _e cab _f the _re engine. _e leader _s usually _en in _e Fire ____ices for _ny years. _e will _ow how _o fight _____rent sorts _f fires. _o, when _e firemen ___ive at a fire, it is always the leader who decides how to fight a fire. He tells each fireman what to do.

It is obvious that certain orthographic cues are also present in the MC-Test. Superficially, there does not seem to be much difference between the C-Test and the MC-Test, since half of every second word is deleted in both cases. However, if one considers the structure of word formation in English, one will easily see the distinction between deleting ~~the~~ first half and deleting the second half of the word, specifically in most content words (i.e., nouns, verbs, adjectives and adverbs), which may normally be inflected. Since all syntactic inflections and most derivational morphemes in English are suffixes, when the first half of the word is deleted, there will be considerable possibility that the remaining will be either part of the base word plus the suffix(es), only the suffix(es), or part of the suffix. As a result, it may be reasonable to conclude that in the MC-Test, more of the syntactic/structural information based on the orthographic cues is present, whereas in the normal C-Test, more of the semantic information is available.

1.3 STATEMENT OF THE PROBLEM

Although it seems that research has been conducted on almost every aspect concerning cloze procedure since its initiation by Taylor in 1953, there still remain many questions that require further clarification. Some research has revealed that the text on which the test is based does influence the validity of the cloze test (e.g., Alderson, 1983). There also seems to be a great deal of controversy over the context which precedes and/or follows a test blank. Whereas some researchers assert that only the immediate context assists the testee in completing the blank, others claim that the passage in its entirety influences the testee's word choice at each blank. (cf., Taylor, 1953, 1972; Oller, 1975; Chihara, Oller, Weaver, & Chavez-Oller, 1977; Rankin & Thomas, 1980; and Shanahan, Kamil, & Tobin, 1982). There are also questions on types of word deletions, frequencies of word deletions, scoring procedure, and how the score is interpreted (Alderson, 1979, 1983; Klein-Braley, 1981, 1983; Foley, 1983; Brown, 1983, 1984; Klein-Braley & Raatz, 1984). And perhaps the most important question of all: ~~If~~ there are still many doubts about the instrument itself, how can one be sure that the technique really reflects ESL proficiency?

The C-Test, a newly-developed version of cloze procedure, initiated by Klein-Braley & Raatz in 1981, has appeared to be a new and challenging measuring technique. Resembling the development of the conventional cloze test at its initial stage, the C-Test was investigated solely by its initiators--Klein-Braley & Raatz, at Duisburg, Germany, and a few other researchers in their group. Perhaps this is because most language educators and testers generally do not have access to studies conducted outside North America, or perhaps simply because the cloze technique has already passed through the stage of challenge. The limited information available, however, suggests that the C-Test has potential as an alternative measuring device that may resolve the problems facing the conventional cloze test.

1.4 PURPOSE OF THE STUDY

It is the aim of this study to provide evidence:

1. to establish the reliability and validity of both the C-Test and the MC-Test;
2. to justify whether different starting points of mutilations applied to the same texts affect the difficulty, reliability, and validity of the C-Test and the MC-Test; and
3. to explore some strategies used by L2 learners in restoring the C-Test and MC-Test passages and to find out their opinions of this type of test.

1.5 RESEARCH QUESTIONS

This study will be guided by the following questions.

1. How reliable and valid are the C-Test and the MC-test as measuring techniques for L2 learners? (Purpose #1)
2. To what extent do different starting points of mutilations applied to the same texts affect the difficulty, reliability, and validity of the C-Test and the MC-Test? (Purpose #2)
3. What are the underlying factors for the C-Test and the MC-Test, as determined by factor analysis? (Purposes #1,2)
4. Is there a relationship between the C-Test/MC-Test scores and the Michigan Test scores

of the L2 learners? (Purpose #1)

5. What are the common types of errors made by L2 learners? To what extent do patterns of errors indicate their reading strategies? (Purpose #3)
6. To discover the extent to which interview data correspond with the results of the error analysis? (Purposes #3,1)

1.6 THEORETICAL RATIONALE

Most researchers of cloze procedure have agreed with Taylor's rationale, namely, that the procedure is based on the Gestalt principle of "closure" or "human ability to fill gaps." The credence of this principle as applied to the deletion of language elements to be reconstructed by observers, however, was once questioned. Anderson (1976, p. 14) pointed out that Weaver (1965) argued that "to relate cloze procedure to the Gestalt notion of closure was a tenuous argument." There was no evidence to support it beyond observing that there was a gap to be filled and that when a subject completed the gap, this was the same process as the subject perceiving the whole for an incomplete pattern. Anderson suggested that evidence presented by Rankin (1965) seemed to support Weaver, for Rankin concluded that "closure of language patterns is primarily a cognitive rather than a perceptual task." However, as Anderson observed, in a later study conducted with Ohnmacht & Kohler, Weaver (1970) appeared to reverse his opinion when they reached the conclusion that "cloze factors and perceptual closure factors are moderately correlated."

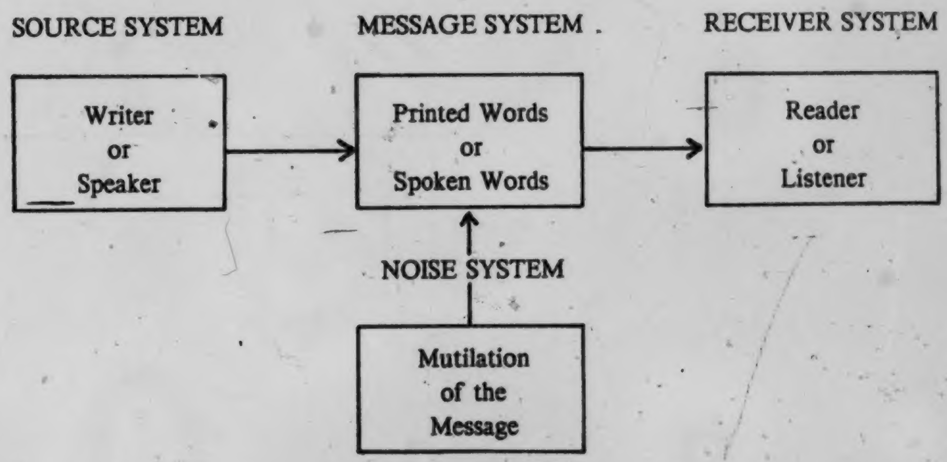
In Taylor's (1953) pioneering paper, cloze procedure was defined as:

> A method of intercepting a message from a "transmitter" (writer or speaker), mutilating its language patterns by deleting parts, and so administering it to "receivers" (readers or listeners) that their attempts to make the patterns whole again potentially yield a considerable number of cloze units. (p. 416)

If one closely examines this definition, one will agree with Anderson (1976) that this rationale of cloze procedure is based on Shannon & Weaver's (1949) "Generalized Communication Model." The model was developed to deal with signal transmission especially as it relates to the telephone. In this case, the source is the different sound vibrations produced by the speaker. A transmitter transforms these into electrical signals which travel over lines (the channel). A receiver then transforms the signals back to sound vibrations and the message reaches its destination, the listener. Interference in the line, which causes distortion of the message, is termed "noise."

The adaptation of the Generalized Communication Model to language communication was made by Osgood (1959), who was Taylor's dissertation supervisor. The three components of Osgood's model, the Representational Model, were termed Source System, Message System, and Receiver System.

As suggested by Anderson (1976), the use of cloze procedure in language communication requires an extension to the model, namely, the addition of a fourth component, "noise." The diagram below shows the extended model, comprising the four components.



A Model for the Language Correspondence of a Source System to a Receiver System.
(From Anderson, 1976, p. 15)

Anderson posits that in this language correspondence model, the transmission and receiving of a message are essentially "coding operations." In cloze procedure, the source which produces a message is the writer or author, and the message, a passage of printed English. Mutilation of the language patterns of the message is "noise" which interrupts the coded message before it is received by the decoder or reader. The reader's task is, therefore, to restore the mutilated passages by using his/her language knowledge and the information available.

Greene (1965) also perceives cloze procedure as a measure of the degree to which the language of a message and the language system of a reader correspond. Greene postulates:

> To the extent that the reader's language facility allows him to utilize the semantic and syntactic information in a passage, he is able to complete the cloze task and communication (comprehension) is assumed to have occurred. A passage that uses an unknown vocabulary or grammatical patterns different from those of the reader is difficult for the reader to complete. Also, a reader with a limited vocabulary or restricted grammatical sophistication has more difficulty in completing a given passage than does a more adequate reader. (p. 213)

In discussing the processing of the various types of cloze passages, Oller (1979) implies that the ability to identify redundancy of the language or what he calls the "expectancy grammar" of the learner is very important. Responses to the cloze items are usually confined by both semantic and syntactic constraints. In the case where only part of the word is deleted, the textual clues of word spaces and unmutilated letters will further constrain the possibilities of responses. Oller, thus, suggests that "we may deduce that the cloze procedure . . . is a method for testing the learner's internalized system of grammatical knowledge" (p. 344).

Regarding the validity of the C-Test, Raatz & Klein-Braley (1981) argue that the test appears to be both theoretically and empirically valid. With respect to theoretical validity, they state:

> The theoretcal basis for the C-Test is identical to that for the classical cloze test. Following Oller and Spolsky we posit an internalized language processing mechanism which can be activated by presenting the examinee with a mutilated text and asking him/her to restore the missing elements. The C-Test is however more in accord with sampling theory since the deletions in the text are, in fact, random, and it has a stronger relationship to "what it means to know a language" since the test scores can be related to the virtually perfect scores obtained by proficient adult native speakers of the language. (p. 134)

In relation to empirical validity, Raatz & Klein-Braley provided evidence to support a claim that the C-Test highly correlates with teacher judgments and other language tests.

The theoretical rationale for using cloze procedure to measure reading comprehension is normally claimed to be based upon current psycholinguistic theories of the reading process. One of the most widely accepted reading models which describe the reading strategies used by readers is that of Goodman (1967). Goodman proposes that readers use graphophonic, syntactic, and semantic information as they engage in the reading act, and that the best readers use the least amount of text information possible. Smith (1978) argues that reading as an activity involves two forms of information: the visual, that is, what is on the printed page, and the nonvisual, that is, the reader's language competence and his background experiences. Through the exchange of these two forms of information, the reader understands, to the best of his abilities, what the author is describing. Therefore, the more nonvisual information one has, the less he needs to rely on the visual information. Although Goodman and Smith use somewhat different explanations for their models, they essentially agree that an efficient reader usually uses a minimum amount of text/visual information.

The rationale for using the C-Test to measure ESL proficiency is, like the rationale for using the conventional cloze test, essentially based upon the psycholinguistic theories mentioned, more specifically, the Goodman Model. In the case of the C-Test, since half of every second word (not the whole word) is deleted, obviously some graphophonic/orthographic cues are still present in each C-Test item. Research reports on conventional cloze procedure confirm that "deleting words more frequently than one out of five creates a test of such difficulty that much discriminatory power is lost" (Oller, 1973, p. 107). However, this problem of difficulty can be resolved by the presence of the graphophonic/orthographic cues left in each C-Test/MC-Test item. Removing half of the word leaves the reader with a fair amount of information even if the deletions occur every second word. The remaining question is, between the syntactic and the semantic cues, which type of cues permits the reader to complete the C-Test/MC-Test passages more easily?

1.7 ASSUMPTIONS UNDERLYING THE C-TEST

Based upon the criticisms made on the conventional cloze test, Klein-Braley & Raatz have proposed a list of six criteria for a new test, namely, the C-Test. In this study, however, Klein-Braley & Raatz's criterion #3 (i.e., "adult native speakers should obtain virtually perfect scores") will not be followed. Under the assumption that adult native speakers differ in terms of their reading abilities and that passages of different levels of difficulty will be used, it is proposed in this study that adult native speakers may not obtain virtually perfect scores. Native speakers' scores will be used only as the criterion to which non-native speakers' scores can be compared. Consequently, the set of criteria for the test in this study will read:

1. it should use several different texts;
2. it should have at least 100 deletions;
3. the deletions should affect a representative sample of the text;
4. only exact scoring should be possible;
5. the test should have high reliability and validity; and
6. native speakers' scores will be used to establish the criteria to which the non-native speakers' scores can be compared.

In addition, the following guidelines will be adopted for deletion.

1. For a one-letter word, the whole word will be deleted.
2. For an even-number-letter word, half of it will be deleted.
3. For an odd-number-letter word, half of the number-plus-one will be deleted.

1.8 DEFINITION OF TERMS

Cloze Test. A language test in which the student has to fill in blanks in a continuous passage. There are many variations on the cloze test but the basic type involves the test constructor selecting a passage, and then deleting every *n*th word.

C-Test. A language test recently developed as an alternative for the conventional cloze test. The test is normally comprised of four to six short passages of roughly equal length. The mutilation of each passage begins in the second sentence by deleting the second half of every second word. The total number of deletions from the passages is a minimum of 100. (Klein-Braley & Raatz, 1984, p. 136)

Graphophonic cue/information. Graphophonic information is comprised of graphic information, phonological information, and phonic information. This set of information includes cues that appear within words, such elements as sound-letter relationships, shapes of words, affixes, known words, and small, familiar words that appear in larger, unfamiliar ones. (Goodman & Niles, 1970, pp. 107-8; Grove, 1981, p. 7). The term is sometimes used in this study in a narrower sense to be equivalent to "orthographic cue."

MC-Test. A modified C-Test. The proposed MC-Test is comparable to the "normal" C-Test. The only difference between the normal C-Test and the MC-Test is that, instead of the second half, the first half of every second word is deleted in the MC-Test.

Orthographic cue/information. This set of cues involves graphic information or lexical spellings.

Semantic cue/information. Semantic information consists of experience, concepts, and vocabulary possessed within the reader. This involves the language used by the author and the reader, on which there must be agreement. If there are differences in the language, such as dialect, there can be problems in the reading process. Also, if the reader's background experiences are weak, or if he has difficulty understanding concepts, then reading will be difficult. (Goodman & Niles, 1970, p. 108; Grove, 1981, p. 7)

Syntactic/structural cue/information. Syntactic information includes sentence patterns, pattern markers, and transformational rules. This set of cues is essentially comprised of structural markers: word order, function words, question markers, verb markers, inflectional endings (such as past markers or markers for person), and punctuation marks. These are also known as "contextual cues." (Goodman & Niles, 1970, p. 108; Grove, 1981, p. 7)

Chapter 2

REVIEW OF RELATED LITERATURE

The purpose of this chapter is to review the pertinent literature in the fields of cloze procedure and reading strategies.The first section (2.1) attempts to summarize the research on the various aspects of cloze procedure relevant to its use as a measure of language ability. Section two (2.2) examines the studies which are related to reading strategies that L2 learners may use in restoring the C-Test/MC-Test passages. The relationship between cloze procedure and reading strategies is also discussed.

2.1 RESEARCH ON CLOZE PROCEDURE

The initial research on cloze procedure and a majority of later studies can be identified as focusing on one of the two uses of cloze: measuring the readability of reading texts/passages or measuring the reading ability of individuals. Taylor (1956) has, in fact, suggested that it seems possible to use cloze method for tapping most of the obvious variables in communication, including "testing the progress of students learning a foreign language" (p. 99). The introduction of using the procedure for testing foreign language proficiency or global language ability, however, was first attempted in 1959 by Carroll, Caron, & Wilds, who employed the cloze technique with high school foreign language students (Oller & Conrad, 1971, p. 184). In this study, the review of literature will focus on the research concerning methodological problems of cloze procedure and attempts made to solve these problems, and the research on cloze procedure as a test of ESL/EFL proficiency.

2.1.1 METHODOLOGICAL CONSIDERATIONS

2.1.1.1 FREQUENCY OF DELETIONS

Most cloze studies have employed two deletion systems: random deletion and systematic/every *n*th deletion. Alderson (1979), however, argues that there are at least three different systems of cloze procedure.

> The first and most general level of definition is "the systematic deletion of words from text," where *systematic* remains undefined. The second definition takes the word *systematic* and divides it into two types of systems: either a random (or, better, pseudo-random) deletion of words, or a rational deletion. A third definition, which is increasingly common in the literature, is the deletion of every fifth word from text (i.e., not just pseudo-random, but a specific deletion frequency).(p. 219)

The random/pseudo-random deletion and the every *n*th deletion are mechanical and objective. Furthermore, no matter how deletion systems are defined, it is obvious that none of them require any special training in test construction. These advantageous characteristics among others such as ease in constructing, administering, and scoring have made the cloze procedure a popular means of testing.

Taylor (1953) compared the random deletion and the every *n*th deletion techniques. He used a random 10% and every 10th word deletion system. It was found that the three passages used were ranked in the same way with significant *F* ratios for between passages, but only the random 10% system resulted in a significant *F* ratio for between subjects. Since these findings were based on short passages and a sample of only twenty-four subjects, Taylor concluded that the two deletion systems would have produced more nearly equivalent results if more deletions had been used. Most researchers, however, find the use of the every *n*th word procedure appealing because of the simplicity.

A variety of word deletion frequencies has been adopted depending on the researcher's purposes or sometimes the length of the passages. It appears that most researchers use either every fifth or every seventh word deletion. Bormuth (1963) and Ruddell (1964) both in fact defined cloze procedure as the deletion of every fifth word.

The question of context sensitivity of the cloze item has been of interest. Taylor's (1953) data on random and every *n*th word deletion systems and varying deletion frequencies showed that all deletion systems (i.e., every 5th, 7th, and 10th word) differentiated between passages but that some conditions were more efficient in discrimination among subjects. His conclusion was: the 35 blanks and every 5th word deletion discriminated better between its 6 subjects than did the 16 blanks, between its 12 subjects (p. 424).

In a later study, Taylor (1972) used 45 prose fragments, systematically drawn from a short story, to explore the relative influences of "preceding," "following," and "surrounding" contexts on cloze items. He reported that the results failed to show any dependent overall difference between the influence of preceding and following contexts on subjects' cloze responses.

Rankin & Thomas's (1980) study proposed to investigate the influence of familiarity upon contextual constraints of the cloze procedure. Subjects were randomly assigned to either a pre- or post-treatment cloze condition. Within each condition, the subjects were also randomly assigned to one of the four deletion patterns: every-fifth, sixth, tenth, or fifteenth. They concluded:

> 1.Familiarity increases cloze scores on tests with all types of words deleted, but the effect of familiarity is greater upon content words than upon function words.
> 2.Greater context [every tenth and every fifteenth as contrasted with every fifth and every sixth patterns] facilitates cloze inferences for content words, but inhibits cloze inferences for function words.
> 3.Content words are more difficult to predict than function words. (p. 52)

The question of how much cloze scores are affected by scrambled sentence order has also been investigated. Carroll (1972) argues:

> Cloze scores are dependent chiefly on what may be called the 'local redundancy' of a passage, that is, the extent to which linguistic cues in the immediate environment (generally in the same sentence) of a missing word tend to supply it. ... Cloze scores are probably more dependent on detection of grammatical than of semantic cues. (p. 189)

This argument appears to imply that cloze items are not sensitive to discourse constraints. Carver (1975-76) seems to agree with Carroll, for Carver posits that scrambling sentence order would probably have little effect on cloze scores. Alderson (1979) systematically varied the size of fixed-deletion ratios (every 6th, 8th, 10th, and 12th word) across passages of differing difficulty and across groups. He found no evidence that increases in context affected cloze scores. Shanahan, Kamil, & Tobin (1982) used fixed-ratio cloze tests from intact passages, randomly scrambled passages, and passages in which the original sentences were extracted and inserted into unrelated text. No significant differences were found in the subjects' performances across these three conditions.

On the contrary, some researchers have found evidence to support that cloze items are sensitive to constraints beyond the level of the sentence. Ramanauskas (1972), for example, found that native English-speaking subjects performed significantly better on the intact cloze passages than on the scrambled cloze passages. Oller (1975) used five orders of approximation to fully sequential prose by a cut and scramble procedure. His findings showed that the native English speaker subjects performed significantly better on cloze items over sequential prose than the same items in a scrambled order. The results also indicated that items in the 50 word chunks were easier than the same items in 25 word chunks and so on.

Chihara, Oller, Weaver, & Chavez-Oller's (1977) study demonstrated that both native and non-native speakers of English performed much better on cloze items in intact texts than they did on the same items in the scrambled texts. The evidence obtained by Yamada (1979) also supported the conclusion that cloze difficulty is significantly affected by scrambling the sentence order. Chavez-Oller, Chihara, Weaver, & Oller (1985), using the original Chihara et al. (1977) data, investigated differences in item difficulty and item discrimination across scrambled and intact passages and across different proficiency levels. They concluded that intersentential context played an important role in facilitating successful closure.

There could be many sources of inconsistency among the results of these studies. One possible source might be the level of difficulty of passages. (It is likely that the easier the passage, the less difference there will be between the cloze scores of the intact text and the scrambled text.) Another probable source could be the methods used in deleting words from passages. When different fixed-deletion ratios (every *n*th word) are used, essentially differing amounts of context will be available.

2.1.1.2 TYPE OF DELETIONS

An assumption for using every *n*th word deletion technique is that all words should be potential deletions and deletions should affect a representative sample of the text (Taylor, 1957). Hence, this method of word deletions is usually used for measuring text readability. However, other types of deletions may be used, depending on the purpose of the test.

Taylor (1957) experimented with three types of deletions: any-word (every *n*th word), hard-word (content words), and easy-word (function words) methods. Based on correlational analyses, he concluded that the any-word method was found to be the best measure for testing reading comprehension and aptitude, and that the hard-word method, the best predictor of prior knowledge of technically worded content.

Greene (1965), in comparing the deletion of every 12th word with the deletion of only content words, found that the latter produced a higher percentage of effective items than did the former(86% vs. 62%). However, Green concluded that the modified procedure produced a more reliable but not more valid measure of adult reading comprehension. In addition, the modified procedure requires more time for test construction and loses objectivity in item selection.

Oller & Inal (1975) have experimented with a cloze test of English prepositions with both native speakers and non-native speakers of English. Every other preposition was deleted from a passage taken from an ESL reader. The correlation between this test and the UCLA ESLPE (UCLA ESL Placement Examination) was .75, and the highest correlation (.68) was found to be with the grammar section of the UCLA ESLPE.

Bachman (1982) attempted to test hypotheses regarding the level of language context measured by cloze tests. Three types of rational deletions were used. They were: *syntactic* cloze (clause-level context), *cohesive* cloze (interclausal/intersentential cohesive context), and *strategic* cloze (parallel patterns of coherence). From the findings, Bachman concludes that cloze tests, using a rational method, can be used to measure textual relationships, namely, cohesion and coherence. The findings appear to support

Clarke's (1979) conclusion that the rational deletion procedure is superior to the mechanical (every *n*th word) deletion.

Markham (1985) investigated the intersentential sensitivity of the rational deletion cloze procedure with college-level students of German. No significant differences were found in performance on sequential versus scrambled cloze tasks in either the exact or the acceptable scoring condition. Markham argues that "concerning the assumed advantages of rational deletion, it should be remembered that some content words may not be involved in maintaining cohesive relationships between sentences" (p. 429). Therefore, he does not agree that merely employing content words as deletions in cloze task will necessarily increase the global sensitivity of the cloze task.

Again, there has been inconsistency among the results of these studies. Among the many possible sources of the inconsistency, one would be the criteria for selecting words to be deleted. It seems that the term "rational" has already been used to be equivalent to "deletion of content words." In that case, the criteria to determine which content words to delete could vary from researcher to researcher. In Markham's (1985) study, for example, the deletion rate in the scrambled passage was allowed to vary from 7 to 9 (p. 425), but no other guidelines of which word to select among the 7th, 8th or 9th, were suggested. Instead, it was the scoring procedure that helped determine the difference in types of word deletions between the rational deletion and the fixed-ratio deletion in this study. Only content-word deletions were scored, and deletions that involved function words were not included (p. 426). As Bachman (1985, p. 549) suggests, criteria are required in order to provide a principled basis for selecting words to be deleted, that is, for designing the contents to be measured by the cloze.

2.1.1.3 NUMBER OF DELETIONS

One area in the construction of cloze tests which has not been investigated thoroughly is the number of cloze items. This variable, in fact, involves such factors as testing time, fatigue, number of passages, and test reliability. Taylor (1956), however, suggests that approximately 50 blanks will be "sufficient to allow the chances of mechanically selecting easy or hard words to cancel out and yield a stable score of the difficulty of a passage or the performance of an individual (p. 48), no matter what types of deletions may be employed. Most researchers have tended to follow this guideline.

Bormuth (1965), however, attempted to investigate systematically the issue of test length in relation to sample size. Bormuth used the same passage to construct tests of five items to tests of 50 items. In addition, for the every fifth word deletion version, five different starting points of deletions were used so that every word in a 250 word segment of the text appeared as a cloze test item. From this study, two sources of errors, the sampling of items and the sampling of subjects, have been taken into account. Bormuth calculated the standard errors for various combinations of numbers of items and numbers of subjects, and proposed a table for estimating the amount of error for a test of *n* items administered to *N* subjects.

Even with Bormuth's table, Anderson (1976) maintains, "there is no clear answer to the question of the number of items to use in cloze tests for a desired measure of reliability." It may, as well, depend on type and/or frequency of deletions made. Bachman (1985, p. 550), however, argues that to be reliable and valid, cloze tests "need not be as long, nor deletions as frequent as has been recommended in the literature." The tests in his study, for example, included only 30 items and used deletion ratios of 1:11. Bachman also observes that his claim is consistant with Rand (1978) who finds that "maximum reliability for cloze tests is achieved with about 25 deletions."

2.1.1.4 SCORING METHODS

One of the more controversial issues in cloze procedure is how subjects' responses should be scored. A variety of methods has been reported for scoring cloze items. The most widely used procedure appears to be exact word scoring. Some researchers, however, have adopted the practice of also giving credit for synonyms. Taylor (1953), for example, compared scores by exact scoring and acceptable scoring and found that the relationship between the cloze scores for the three passages remained essentially the same. Miller & Coleman (1967) contrasted exact word scores with partially acceptable scoring (exact replacement = 3, synonym = 2, grammatical class = 1). They obtained an almost perfect correlation of .99.

There have been other methods of scoring the cloze test. In a cloze symposium (Greene, et al., 1967, p. 123-124), Taylor suggested four different cirteria for scoring cloze responses. The four criteria were: 1) criteria scoring--the experimenter makes a decision about the criteria; 2) communality, or inter-agreement scoring--attention is paid to the inter-agreement among subjects' responses; 3) latency scoring--the focus is on how many seconds it takes the subject to produce the only one acceptable answer; and 4) measure of the subject's gambling instinct--the subjects are told that they will be punished if they guess wrong, and the experimenter counts the number of times each subject dares to guess. Taylor suggested that for the fourth type of scoring, the passage must be a very difficult one so that no one could guess the word. Unfortunately, Taylor did not mention the value of this type of scoring.

Taylor & Waldman (1969) proposed to specify and differentiate the uses of three approaches of scoring: the "right-wrong" scoring, the "time" or "latency" scoring, and the "subject-interagreement" or "focus" scoring. One interesting finding was "good-to-poor latencies were significantly correlated with good-to-poor performances on the usual tests, ratings, and rankings of 'intelligence' (verbal and non-verbal), short- and long-term memory, reasoning, and environmental orientation, but, surprisingly, *not* with simple reaction time ('Press the key as soon as the light goes on')" (pp. 245-246). Based on the response interagreement data, Taylor & Waldman concluded:

> Counts of numbers of blanks on which subjects interagree at , ..., might discriminate between the comprensibility levels of different samples of prose just as well, or better than, the counts of right-wrong scoring. Also , the number of blanks in which an individual's response agrees (or disagrees) with the majority's responses (whether they are "right" or not) might index something about that individual's ability, personality, or social or educational background.
>
> Further, the response-interagreement concept might be applied to the construction of prose passages which would closely reflect the average vocabulary habits of a particular population. (p. 257)

Alderson (1979) developed five scoring procedures to use in his study. These procedures included: 1)the exact word procedure, 2) the semantically acceptable procedure, 3) the same form class procedure, 4) the acceptable form class, same grammatical function procedure, and 5) the grammatically correct procedure. It was found that "different scoring procedures nearly always result in significantly different mean scores." However, a high level of intercorrelation among the five scoring procedures was obtained, with the semantically acceptable procedure correlating as much to the grammatical procedures as to the exact word procedure. Nevertheless, Alderson noted that:

> The results show that the existence of a high intercorrelation between scoring procedures is not a sufficient criterion to judge one cloze scoring procedure against another, that other evidence needs to be taken into account, and that, indeed, the other evidence may be more important and lead to different conclusions. (p. 205)

In general, it is more convenient to use the exact scoring rather than the more complicated systems if no special interest in differing degrees of acceptability or needs for tapping any specific characteristics of the cloze items are considered.

2.1.2 CLOZE PROCEDURE AS A MEASURE OF ESL/EFL PROFICIENCY

Since the cloze technique has been shown to be a useful measure for testing the language skills of native speakers, it would be expected to be useful with non-natives as well. Johnson (1981), however, argues that the use of cloze procedure with non-native speakers involves a number of assumptions. One assumption that has to be challenged is that the constraints under which native speakers and non-native speakers operate are the same. Another assumption is that the second language is acquired for the same purposes as the mother tongue. Johnson goes on to use Carroll's (1971, p. 177) assertation to suggest that it is not possible to assume that "the second language speaker uses that language in ways which are directly comparable to the first language speaker." The experience of learning, primarily in the classroom (as contrasted with "acquisition" in Krashen's term), is very different from the experience of learning/acquiring a language naturally as part of the general process of socialization and maturation.

Freeland (1979) points out:

> Certain assumptions about cloze tests need reexamining. ... And the practice of expressing learners' scores as a ratio of native speakers' is suspect, since the relationship between native and non-native scores can fluctuate in unknown ways. (p. 6)

Freeland argues that research studies support that for non-native speakers, text type can be an important source of variation. Native speakers, on the other hand, have the competence that enables them to identify different registers and styles; therefore, text type is not normally a source of variation in their cloze scores.

Oller (1973a) appears to come to a similar conclusion: "There is little if any reason to assume that conclusions from research with native speakers can validly be generalized to the case of non-native speakers" (p. 107). With this underlying assumption, Oller argues that "when dealing with non-native speakers there is something counter-intuitive about requiring the exact word." This argument has been well supported by research evidence which concludes that with non-native speakers the acceptable word scoring method is superior to the exact scoring. It has also been noted that exact scoring often makes tests too difficult for non-native speakers and that exact word replacement often requires insights which may not be regarded as language skills (Oller, 1972).

Many researchers have investigated the use of cloze procedure as a measure of second or foreign language proficiency. Among these several studies, a few will be mentioned here.

Darnell (1968) used a scoring method which was based on the criterion determined by the distribution of response frequencies to each item as answered by native speakers. A special computer program was employed to calculate the complicated data. It was found that a correlation of .83 was obtained between his technique and the TOEFL.

Oller & Conrad (1971) conducted an experiment to determine the discriminative power of a cloze test and its validity as a device for measuring ESL skills. The exact word scoring was used. The cloze test was found to highly correlate with the UCLA ESL Placement Examination, Form 2C (.88), and correlate best with the dictation section (.82). Oller concluded that the cloze technique is a very promising device for measuring ESL proficiency. Further, the cloze procedure along with other operational tests of global language skills may provide very useful information concerning the top cutoff levels for ESL placement examinations."

Oller, Bowen, Dien, & Mason (1972) made an extensive comparison of native and non-native performance on cloze tests. Cloze tests were constructed in English, Thai, and Vietnamese. The subjects were the native speakers of Thai and the native speakers of Vietnamese. The subjects who were native speakers of Thai and Vietnamese took the tests in their native languages and took comparable tests in English. Response frequencies for native speakers of English, Thai, and Vietnamese were systematically compared with response frequencies for non-native speakers of English. The researchers found evidence to suggest that translating a cloze passage from one language into another could produce two tests of approximately equivalent difficulty for similar groups of native speakers of the respective languages.

Alderson (1979) reported on a series of experiments conducted on the cloze procedure where the variables of text difficulty, scoring procedure, and deletion frequency were systematically varied. The main purpose of these studies was to investigate the effects of the variations examined on the relationship of the cloze test to measures of proficiency in English as a foreign language. The following findings were reported: 1) the difficult text correlated considerably higher with the Grammar, Vocabulary, and Reading Comprehension sections on the ELBA test (a test of proficiency in English as a foreign language); 2) the semantically acceptable word scoring almost always correlated higher than the exact word scoring procedure; and 3) the changing of the deletion rate on any given text usually resulted in a difference in correlation coefficients. Alderson concluded that "the cloze procedure is not a unitary technique, since it results in tests which are markedly different; different tests give unpredictably different measures, at least of EFL proficiency" (pp. 225-256). He thus warned testers and teachers not to assume that the procedure would produce automatically valid tests of EFL proficiency.

2.1.3 THE MODIFIED CLOZE PROCEDURE

Research on cloze procedure has continued for more than three decades since it was initiated by Taylor in 1953. The majority of studies has shown evidence to suggest that cloze procedure is a reliable and valid device for measuring passage readability, reading comprehension, and general second/foreign language proficiency. Like most other techniques, the cloze procedure has not been without criticism. A number of criticisms concerning the reliability, validity, and practicality have been made, and have stimulated many researchers to attempt to modify the procedure to overcome the perceived weaknesses. In this study, a few of such attempts will be reviewed.

2.1.3.1 MULTIPLE-CHOICE CLOZE TESTS

Some critics have proposed that cloze procedure could be improved by adopting a multiple-choice format. Porter (1976), for example, argues that in the conventional cloze procedure, the productive language aspects are essentially involved, since the student has to write words to fill the gaps. He suggests that for testing reading comprehension in ESL, a multiple-choice format would reduce the emphasis on the productive process, and facilitate the student in focussing on the comprehension of the passage. Another claim is that writing the responses distracts the reader from the reading task, and a multiple-choice format is a solution to this problem (Ozete, 1977). The most obvious advantage of a multiple-choice format, as Jonz (1976) argues, is that it can be objectively scored.

In modifying a cloze test to a multiple-choice format, one crucial issue is the problem of selecting distractors. The following general guideline for distractor selection for a multiple-choice cloze test has been suggested by Porter (1976).

> The distractors chosen may be varied according to the depth of linguistic attainment and fineness of stylistic discrimination of the student. At an elementary level, the incorrect alternatives may all be quite unsuitable, both in meaning and grammaticality. At a more advanced level they may be grammatically suitable but not semantically appropriate, or *vice versa* . At a fairly high level, distractors may all be grammatically suitable, the student being forced to make quite sophisticated choices among vocabulary items, etc. Following this line of thought further one might offer distractors which are all possible but which vary in probability in the particular register of the passage.

(p.154)

This guideline, in fact, is useful and applies to any multiple-choice format test. In practice, however, it is not always possible to follow this approach, especially, when the every *n*th word deletion procedure is employed. From personal experience of the author of this study, when it comes to a very easy item such as a preposition or many other function words, it is often difficult enough to find even one distractor which is not obviously unsuitable.

Jonz (1976) attempted an approach to selecting distractors for a multiple-choice cloze test of general language proficiency by using three most frequent unacceptable responses from the open-ended cloze test as distractors. He conducted an item analysis and discarded the items which did not meet the criteria by replacing the deleted words in the passage. Jonz reported that the final version had a reliability of .76 and correlated well with class placement and composition scores. A similar approach was in fact investigated by Cranney (1972-73). In the Cranney study, however, the multiple-choice cloze responses were found to be less reliable than the free responses.

Ozete (1977) used a complicated random sampling in selecting distractors for a two-option multiple-choice cloze test for use in testing low-level reading comprehension ability in second language. The two 28-item cloze tests in Spanish were administered to students of Spanish as a second language. Test reliabilities were found to be .57 and .68.

Scholz & Scholz (1981) compared the standard open-ended cloze test with four types of multiple-choice cloze tests that differed in the method by which distractors were selected. Two passages were used, one involving science as an academic subject, and the other involving science as a topic of popular interest. In addition to an open-ended cloze test, four multiple-choice (MC) cloze tests were developed: the interlingual MC cloze test, the revised interlingual MC cloze test, the intralingual MC cloze test, and the teacher-made MC cloze test. They found that the open-ended and the teacher-made multiple-choice cloze tests obtained the highest validity coefficients. They concluded that the open-ended cloze test appeared to be a slightly more valid measure of English proficiency than multiple-choice cloze tests for the academic passage. However, the intralingual and the teacher-made MC cloze tests appeared to be as valid as the open-ended cloze test for the popular passage.

The studies on using a multiple-choice format of the cloze tests appear to yield quite consistent results. Generally, the multiple-choice cloze test tends to be less reliable and valid than the standard open-ended cloze test. The main source underlying the problem in the multiple-choice cloze test is obviously the procedure in selecting distractors.

2.1.3.2 MATCHING CLOZE TESTS

Another attempt to modify the cloze procedure is the matching cloze test. Propst & Baldauf (1979, 1981) propose the matching cloze test as a holistic reading measure which, they claim, requires use of the skills needed in actual reading situations. The process for constructing, administering, and scoring matching cloze tests, as Propst & Baldauf suggest, is as follows:

> Matching cloze tests are constructed by selecting segments of text of approximately twenty-five to thirty-five words in length and deleting five words, usually every fifth. The five deleted words, which are replaced by blanks, are placed in a list to the right of the passage. The test taker's task is to write the correct word, selecting from the words in the list on the right, in each of the blanks. For advanced readers, a sixth word is added to the list at the right of the passage. A test consists of fifty blanks or ten segments of text, each segment containing five blanks. (Propst & Baldauf, 1981, pp. 85-86)

Propst & Baldauf claim that the use of matching cloze tests to measure the reading comprehension skills of beginning ESL students conforms to current psycholinguistic views of the reading process, and more specifically the Goodman (1970) model. According to Goodman, good readers are constantly involved in an editing process as they read, asking themselves whether what they are reading makes sense in terms of meaning and in terms of the syntactic rules of the language. Propst & Baldauf argue that the matching cloze test requires the student to use this editing process. The additional information provided in the form of word choices will facilitate beginning ESL students (especially those who have not developed effective reading strategies even in their own language) in processing the text.

Matching cloze tests, as far as literature is concerned, have not been widely investigated. Research into this technique has been limited to only that of Propst & Baldauf and few unpublished students projects at Townsville, Australia. Although the technique is somewhat similar to the multiple-choice cloze test, that is, supplying choices for students, the matching cloze test is much easier to construct than the latter. The matching cloze test can reduce the problem of selecting the alternatives confronting the multiple-choice cloze test. If Propst & Baldauf's suggestions are strictly followed, they will be suitable for only beginning ESL students, which is the aimed population of the investigators. It seems, however, that this format of cloze procedure may also be effectively used with more advanced students if the technique is modified. For example, if longer passages rather than passages of five deletions each are used and more word-choices are added, the test might be used with the more advanced levels. This suggestion, of course, needs further investigations to provide evidence to support it.

2.1.3.3 WELL-TAILORED CLOZE TEST

Brown (1984), in discussing how a cloze test can be made to fit a given sample, has proposed that cloze tests, like any other language tests, should be pretested. Among the three methods suggested, the most interesting one is the "well-tailored" cloze test. The technique has been described as follows:

> Five different, but non-overlapping, every 7th word deletion pattern versions of one passage (50 items each) were administered to random samples of a group of Chinese students who had a very narrow range of talent. Analysis of the results produced item difficulty and discrimination indices for a pool of 250 possible items. From these items the *best* 50 were selected. In other words, those which had item difficulty levels most closely approximating .50 and the highest discrimination indices were chosen. One restriction was placed on this selection process. The distance between items on the final version was to be no less than five words and no more than nine with an average of seven words. (pp. 117-118)

This final version of the test was readministed to the same group of subjects after six months (to avoid practice effect) and found to be much more reliable than any of the original versions with the same group. Brown concludes that "a cloze test can be tailored to fit a given group in much the same way.that discrete-point tests have traditionally been developed" (p. 118). However, as Brown also points out, the same precision as in the case of the discrete point tests may not be obtained due to the difference in the context provided in the various versions involved.

2.1.3.4 C-TESTS

One of the latest alternatives developed to resolve the problems and questions confronting the conventional cloze test is the C-Test. This modified form of cloze procedure was developed at the University of Duisburg, Germany by Raatz & Klein-Braley in 1981. The research in this area appears to be limited to only that of these two investigators and a few other researchers in their group; however, the results to date have been quite promising. A summary of the investigations will be presented.

In their first report on the development of the C-Test, Raatz & Klein-Braley (1981) described the test development procedure and the findings of their pilot studies with both L1 and L2 samples. They concluded that the results from both data were promising and that the C-Test principles appeared to conform more with "random noise" as defined by information theory than the conventional cloze test.

Raatz (1984) attempted to establish the factorial validity of a German C-Test. His subjects consisted of 75 fifth grade students from three different types of German secondary schools. The C-Test was found to correlate most highly with the orthography test, the grades for German and grammar, and the subtest measuring the ability to structure written material. The C-Test also correlated moderately high (between .4 and .6) with the English grade, the vocabularyand reading comprehension tests, and the tests measuring intelligence. However, no relationship was found between the C-Test and the concentration test. Between the two factors which emerged, the C-Test was found to load highly on the first factor which was interpreted as a general language proficiency factor.

In 1984, Cohen, Segal, & Weiss Bar-Siman-tov summarized the results of two studies (Segal, 1983; and Weiss, 1983). These studies were conducted to investigate the reliability and validity of the Hebrew L2 C-Test. They concluded that the Hebrew C-Test was both reliable and valid as a measure of general language ability. However, they suggested that since the first half of the word was given, the C-Test might encourage word-level processing rather than high-level processing (i.e., processing at the connected discourse level).

In a survey of the C-Test research, Klein-Braley & Raatz (1984) presented a summary of the studies conducted up to that period of time. They concluded that the C-Test was experimented with three different groups: L1, L2, and FL. Furthermore, investigations were performed in six languages: English, French, German, Spanish, Hebrew, and Turkish. The results were all very promising, even in the last two which were non-Indo-European languages.

In a later report on the construct validation of the C-Test, Klein-Braley (1985) presented a discussion of test authenticity and validity. Four hypotheses for C-Test construct validation were proposed.

> 1. If the same C-Test is administered to subjects at different stages of language development, then the C-Test scores will become successively higher as the subjects become more proficient in the language. (p. 84)
> 2. Subjects learning a language 'naturally' will exhibit similar behaviour on C-Tests in that language. (p. 86)
> 3. If texts have an inherent 'C-Test processing difficulty' which is independent of the subject groups involved then it will be possible to discover characteristics of the texts which can be used to predict the rank order of difficulty of texts, possibly even the actually empirical difficulty levels, for specific subject groups. (p. 89)
> 4. Learners with more efficient language processing strategies will make higher scores on C-Tests. (p. 97)

Klein-Braley provided theoretical considerations and empirical evidence to support these hypotheses and concluded that there has been "sufficient evidence to substantiate the claim that C-Tests are authentic tests of the construct of general language proficiency" (p. 101).

2.2 RESEARCH ON READING STRATEGIES

This section proposes to identify research studies which are pertinent to understanding the reading strategies that L2 learners may use in completing the C-Test/MC-Test passages. The first part (2.2.1) suggests the models of reading processes that seem to be appropriate for the purpose of this study. In the second part (2.2.2), common reading strategies derived from research investigations are outlined. The last part (2.2.3) attempts to review recent studies in second language reading processes and models.

2.2.1 THE READING PROCESS

A number of reading models have been proposed based on current psycholinguistic theories. Models of reading are important, for they "represent a set of assumptions about what happens when people read, that is, about the ways that readers go about deriving meaning from a printed text and the relative importance of various aspects of the reading process" (Devine, 1983, p. 95). Despite the various theoretical models[1] purporting to explain how meaning is derived from written language by the reader, most of them, as Cziko (1980) suggests, can be conveniently categorized into three distinct groups: bottom-up, top-down, and interactive views.

> A bottom-up view of reading describes reading as a one-way flow of information starting with the visual (graphic) input and proceeding through a series of progressively higher-order processing stages until meaning is derived. ...
>
> A top-down view of reading emphasizes the role of higher cognitive processes that generate meaning hypotheses based primarily on contextual information. Proponents of this view of reading consider the reader actively involved in deriving meaning directly from the page, without the use of an intermediate speech code, using knowledge of the syntactic and semantic systems of the language, the discourse constraints of the passage, and extra-linguistic knowledge of topics related to the passage. ...
>
> Finally, an interactive view of reading attempts to describe how the reader uses and integrates both graphic and contextual information in extracting meaning from written language. Theories within this group combine and expand upon features of both bottom-up and top-down views of reading and do this within an information processing analysis of language comprehension. (Cziko, 1980, pp. 101-102)

---

[1] Singer & Ruddell's (Eds., 1985) Section Three provides a comprehensive discussion of current theoretical models of reading processes.

Other researchers (Harste & Burke, 1977; Rhodes, 1979; and Devine, 1983) identify these models based on "the language focus and unit emphasized." These models are: 1) a decoding or sound-centered model, 2) a skills or word-centered model, and 3) a meaning-centered model. These latter classifications will be adopted to identify ESL readers' strategies in the present study.

2.2.2 READING STRATEGIES

Many studies on reading strategies have focused on the differences between good readers and poor readers. This type of research has been referred to as "good and poor reader research." Aulls (1981) has identified four common research designs of which the first three have typically been used in good and poor reader studies. These four research designs are:

> 1. Studies which contrast good and poor readers of the same age on specific performance tasks in order to determine differences in skill knowledge. For example, knowledge of words, syntactic cues and main idea units, etc.
> 2. Studies which compare good and poor readers of the same age on one or more processing strategies. For example, prediction of form class, self-correction of mistakes, higher order word grouping of information, inferencing, etc.
> 3. Studies which correlate knowledge and/or strategies observed among good and poor readers in order to study the differences between group correction patterns.
> 4. Studies which provide short training tasks to good and poor readers (using one or more age groups) where training is expected to enhance poor reader performance but not good reader performance given specific hypotheses regarding differences in access to strategies or skills. (p. 87)

Cooper & Petrosky (1976), in their article *A Psycholinguistic View of the Fluent Reading Process*, have presented a description of a psycholinguistic model of the reading strategies derived from the model. These strategies are:

> 1. The reader discovers the distinctive features in letters, words, and meaning. ...
> 2. The reader takes chances--risks errors--in order to learn about printed text and to predict meaning. ...
> 3. The reader reads to identify meaning rather than to identify letters or words. ...
> 4. The reader guesses from context at unfamiliar words, or else just skips them. ...
> 5. The reader takes an active role, bringing to bear his or her knowledge of the world and of the particular topic in the text. ...
> 6. The reader reads as though he or she expects the text to make sense. ...
> 7. The reader makes use of redundancies--orthographic, syntactic, and semantic--to reduce uncertainty about meaning. ...
> 8. The reader maintains enough speed to overcome the limitations of the visual processing and memory systems. ...
> 9. The reader shifts approaches for special materials. ...

10. The reader shifts approaches depending on the purpose. (pp. 191-195)

Goodman, Burke, & Sherman (1980) argue that since the language, thoughts, and meanings of an author are different from those of the readers, reading can never be an exact process and readers can never be sure that they have discovered the meaning the author intended. Thus, the readers will have to use a number of strategies to solve this major problem. Goodman, Burke, & Sherman posit that the significant strategies involved in this process are: predicting, confirming, and integrating. They claim that these strategies are used by all readers, regardless of their reading proficiency. In support of the interactive-compensatory model which purports to explain individual differences in reading, Stanovich (1980) argues that "given a deficit in a particular process, the reader would rely more on other knowledge sources, *regardless* of their level"(p. 34). This thus implies that when facing a difficult problem, not only the good reader but the poor reader as well would try to make use of all the information available to solve it.

2.2.3 STUDIES IN SECOND LANGUAGE READING

Many L2 reading researchers agree that L2 reading is not much different from L1 reading. Specifically, Goodman's model of reading has often been adopted to explain the L2 reading process.

Coady (1979) attempted to relate Goodman's (1967, 1971) and Smith's (1971) approaches to the ESL reading process. He views ESL reading as interaction among three essential factors: high-level conceptual abilities, background knowledge, and process strategies. Coady argues that conceptual abilities are necessary in reading acquisition, especially in adult foreign students. The importance of background knowledge is evident, since students with a Western background of some kind normally learn English faster than others. He considers process strategies as important subcomponents of reading ability. These process strategies include: grapheme-morphophoneme correspondences, syllable-morpheme information, syntactic information (deep and surface), lexical meaning and contextual meaning, cognitive strategies, and affective mobilizers.

Coady claims that:

> ... ESL students typically begin by attending to more concrete process strategies such as phoneme-grapheme correspondences and word meaning. But gradually the student learns to take advantage of the more abstract process strategies such as context and syntax. This change takes place as the reader decides that a particular skill or combination of skills is not working as well in deriving meaning (comprehending) as another combination might. This change or shift, therefore, signifies awareness on the part of the reader that these skills are also strategies for successful comprehension--hence the choice of the term process strategies. (1979, p. 8)

He further points out that many process strategies are primarily related to knowledge of the target language. However, some ESL students whose proficiency in English is high and yet read slowly and without much comprehension, obviously have a reading problem and not a language problem.

He concludes that in learning to read in a second language, the target language must be mastered. In addition, reading ability in the native language may transfer automatically to reading in the second language. However, teachers will have to teach many reading skills which should have been learned in the first language. As well, aspects of the process strategies which are language- and culture-specific must be taught. On discussing transfer from native to foreign language reading, Ulijn (1984) provided evidence to support that the transfer (in a positive or negative way) is more salient on the lexicon than on the syntax (p. 70).

Cooper (1984) attempted to compare what he calls practised and unpractised readers and examine the linguistic features that might cause problems in L2 reading. It was found that the unpractised readers were inferior in using the linguistic clues in the larger context to determine meaning, more severely disadvantaged by their poor knowledge of vocabulary, and displayed weaknesses in all the other main areas investigated.

Berman (1984) investigated the effect of the syntactic components on foreign language reading. She argues that "efficient FL readers must rely--in part, though not exclusively--on syntactic devices to get at text meaning" (p. 139). She also claims that "successful reading, like all linguistic decoding, requires the reader to extract the semantic gist of the language material confronting him" (p. 140). She found that of the two groups, readers of the simplified text did consistently better on all types of questions. Readers of the original version were found to make relatively more intrasentential errors than general gist errors. Berman thus suggests that syntax is more important at the sentence level.

Levenston, Nir, & Blum-Kulka (1984) proposed the use of cloze techniques in testing discourse. They argue that "in discourse cloze testing, where deletions are non-random and consist solely of cohesion-markers, no account is taken of the student's ability to interpret relationships at the micro-level" (p. 207). Therefore, an understanding of the macro-structure could be measured by a discourse cloze.

Hosenfeld (1977) conducted a study to investigate the reading strategies of successful and nonsuccessful L2 readers. Two hundred and ten L2 learners were administered the MLA-Cooperative Test of Reading Proficiency and 20 students who scored high and 20 students who scored low were selected for the investigation. Interviews with these 40 subjects were conducted using the "think aloud" technique, and the following interviewing principles were used:

(1) Try to obtain a complete description of a student's reading strategy. The reading strategy is viewed as comprised of two categories of operations: "main meaning line" (sentence level) and "word-solving strategies" (word level).

(2) Elicit an introspective rather than a retrospective description of a student's reading strategy. In other words, encourage the student to describe his/her operations as he/she reads rather than after he/she has read a sentence, several sentences, or an entire task.

(3) Use indirect rather than direct questions.

Hosenfeld concluded that:

> A distinguishing characteristic of successful and nonsuccessful readers is the priority system of their word-solving strategies: while looking up words in a glossary is a nonsuccessful reader's first and most frequent response, it is a successful reader's last and most infrequent response to unknown words. (p. 121)

Cziko (1980) compared the French oral reading errors of intermediate and advanced English speaking learners with the errors of native French speaking students. Each subject was tested individually by the researchers. Subjects were instructed to read each of the two texts aloud and to try to understand it since comprehension questions would be asked after each reading. Each subject's test session was tape-recorded, scored, and then analyzed.

Cziko concluded that the results of the study indicated that reading strategies were related to the reader's level of competence in the language. While native speakers and advanced learners of French appeared to use an interactive strategy of integrating both graphic and contextual information in reading French, readers with less competence in French would tend to use a "bottom-up" strategy, relying primarily on graphic information.

Devine (1983) investigated the theoretical orientation (internalized model of reading) and reading performance of 20 beginning/low intermediate adult ESL students in East Lansing, Michigan. Three sets of data were collected from each subject: an oral interview, a sample of oral reading, and a retelling (summary) of the oral reading. The reading interview was focused on the subjects' responses to questions designed to uncover general attitudes about reading, idea about what constitutes good/effective reading, notions about the relative importance of various aspects of reading, and information about the way individual readers use the various cues available in a printed text. Using models proposed by Harste & Burke (1977), the subjects were classified (according to the language units they professed to focus on or indicated they considered important to effective reading) as sound-, word-, or meaning-oriented. The data provided by the reading interview were then compared to the subjects' oral reading behavior and to their retellings in order to assess the impact of the readers' internalized models on reading performance.

It was found that 19 out of 20 subjects could articulate their theoretical orientations toward reading unambiguously enough to be identified: six as sound-centered, seven as word-centered, and the remaining six as meaning-centered. There was a correspondence between the subjects' models of reading and the types of information (graphic/sound, syntactic, or semantic) that the readers focused on in oral reading. In general, a relationship was found between the internalized model of reading and the success of the reader in comprehending text material. The meaning-centered readers demonstrated good to excellent comprehension whereas the sound-centered readers were found to have either poor or very poor comprehension. However, the comprehension pattern of the word-centered readers appeared a little inconsistent among these readers.

Block (1986) used a "think aloud" technique to elicit comprehension strategies of nine poor readers. These readers were college-level students (three native speakers of Spanish, three native speakers of Chinese, and three native speakers of English) enrolled in remedial reading classes. Block concluded that readers of different language backgrounds did not appear to use strategies that were different.

The review of literature seems to suggest that the strategies that L2 learners use in completing cloze passages may be accounted for by their type of readers. However, little or no attention has been focused on the relationship between L2 learners' strategies in restoring cloze passages and reader types. This study will attempt to draw relationships between strategies employed in reading cloze passages and normal reading strategies.

Chapter 3

PROCEDURE

3.1 TEST CONSTRUCTION

3.1.1 SELECTION OF PASSAGES

Six short passages of different readability levels were selected from ESL textbooks and graded readers. The content of each passage was of a general theme, which would not require any special knowledge or terminology. The readability levels of passages were determined by using Fry's (1977) Readability Scale. The following are the sources and readability levels of the six passages.[2]

Passage 1: *New Practice Readers Book G* (Grover & Anderson, 1961, p. 44); grade 4/5.

Passage 2: *Elementary Stories for Reproduction* (Hill, 1967, p. 16); grade 5/6.

Passage 3: *New Practice Readers Book G* (Grover & Anderson, 1961, p. 4); grade 7.

Passage 4: *New Practice Readers Book F* (Grover, Kinkead, & Anderson, 1962, p. 46); grade 8.

Passage 5: *Reading Faster: A Drill Book* (Fry, 1964, p. 43); grade 9.

Passage 6: *New Practice Readers Book F* (Grover, Kinkead, & Anderson, 1962, p. 78); grade 10.

Based on the results of the pilot study, Passages 1 and 5 were omitted, resulting in four subtests for each form of test for the main study.

3.1.2 MUTILATION

Four different forms of tests were constructed from the four passages chosen. These were two forms of the C-Test (C-Test 1 and C-Test 2) and two forms of the MC-Test

---

[2] The passage used as examples for the C-Test and the MC-Test in the test instructions was adapted from Ruddell, Taylor, & Adams (1978, pp. 4-5).

(MC-Test 1 and MC-Test 2).

C-Test 1. In C-Test 1, each of the four passages was mutilated by deleting the second half of every second word, starting from the second word of the second sentence. The deleted part of the word was replaced by an appropriate blank to correspond with the number of the missing letters. There were 25 mutilations in each passage, totalling 100 mutilations or C-Test items for the whole test.

C-Test 2. C-Test 2 was constructed in the same manner from the same four passages. The only difference from C-Test 1 was that the deletion started from the third word, instead of the second word, of the second sentence. There was also a total of 100 C-Test items on this form.

MC-Test 1. MC-Test 1 was constructed in the same manner as in C-Test 1 in that the deletion in each of the four passages started from the second word of the second sentence. However, the first half, instead of the second half, of every second word was deleted. There was a total of 100 items in this form as well.

MC-Test 2. MC-Test 2 was constructed in the samemanner as in MC-Test 1, from the same four passages. The only difference from MC-Test 1, was that the deletion started from the third word, instead of the second word, of the second sentence. The total number of mutilations was also 100.

3.1.3 SCORING METHOD

Exact word scoring was used. Each mutilation was considered one item. Every correct response was given one point, while an incorrect answer or non-response was given zero.

3.2 CRITERION MEASURES

1. A non-secure form of the Michigan Test of English Language Proficiency (The Michigan Test), specifically, Form B, which is a standardized test, was used as a criterion measure of ESL (L2) subjects' English language proficiency.

2. C-Test/MC-Test scores achieved by native English speaker (L1) subjects were used to establish the criteria to which L2 subjects' scores could be compared.

3.3 SAMPLES OF THE STUDY

The L1 sample consisted of 389 students from two junior high schools and one composite high school in the Edmonton Public School System. Distribution of grade levels and test forms is shown in Table 3.1.

The L2 sample was comprised of 104 students enrolled in ESL courses at an adult education center in Edmonton. These students had had formal English instruction in Canada for between less than a year and six years, with a mean of 1.5 years. Length of their stay in Canada ranged from less than a year to 11 years, with a mean length of stay of 3.4 years. ~~Distribution of~~ their language backgrounds and test forms is shown in Table 3.2.

3.4 DATA COLLECTION

1. The four forms of tests were randomly administere~~d to the~~ L1 subjects by their language arts teachers during the week of May 12-16, 1986. Each subject was administered one test form. Written directions by the researcher were given to these teachers so that test administration procedures were the same for all groups[3] (see Appendix A).

2. The Michigan Test was administered to the L2 subjects on May 21, 1986, and the same four forms of the C-Test and the MC-Test were administered to them on the following day. The administrations of these tests were performed by the students' instructors who had been requested to follow strictly the standard administration for the Michigan Test and the written directions by the researcher for the C-Test/MC-Test.

3. Of the 104 L2 subjects, 28 were individually interviewed by the researcher to reveal the reading strategies they used in restoring the C-Test/MC-Test passages. These subjects were randomly chosen to equally represent the high proficiency group (HPG; the top quarter)

---

[3] Note that the C-Test/MC-Test were considered as "power" tests; thus, there was no time limit for completing the tasks.

Table 3.1

Distribution of Grade Levels and Test Forms for L1 Sample

Grade Level	Test Form				Total	
	C1	MC1	C2	MC2		
Grade 8	26	22	26	19	93	(23.9%)
Grade 9	25	24	24	21	94	(24.2%)
Grade 10	25	24	25	24	98	(25.2%)
Grade 11	26	25	27	26	104	(26.7%)
Total	102	95	102	90	389	
	(26.2%)	(24.4%)	(26.2%)	(23.1%)		

Table 3.2

Distribution of Language Backgrounds and Test Forms for L2 Sample

Language	Test Form				Total	
	C1	MC1	C2	MC2		
Vietnamese	11	7	7	11	36	(34.6%)
Chinese	7	11	7	9	34	(32.7%)
Spanish	5	5	3	3	16	(15.4%)
Polish	0	1	4	2	7	(6.7%)
Others	3	2	5	1	11	(10.6%)
Total	26	26	26	26	104	
	(25.0%)	(25.0%)	(25.0%)	(25.0%)		

Note: Other languages included Amhric (Ethiopia), Cambodian, Czech, Gujarati (India), Hungarian, Lao, Persian, Portuguese, Punjabi (India), and Rumanian. There was one subject from each language except for Persian which had two.

and the low proficiency group (LPG; the bottom quarter). The interviews were completed during the week of May 26-29, 1986.

3.5 DATA ANALYSIS

1. The C-Test/MC-Test results from L1 and L2 data were separately analyzed by using TEST23 (LERTAP), a computer program supplied by the Division of Educational Research Services (DERS), The University of Alberta. The statistics obtained from these analyses were:

    a. Item difficulty;

    b. Item discrimination;

    c. Hoyt estimate of reliability; and

    d. Cronbach's alpha.

2. In order to ascertain how the various items clustered, factor analysis was performed on the results of each subtest from L1 data, using FACT20, a computer program developed by DERS.

3. Correlation coefficients between the Michigan Test scores and the C-Test/MC-Test scores of the L2 subjects were computed by using DEST02, a computer program developed by DERS.

4. Error analysis was conducted on the C-Test/MC-Test responses of the 28 interviewees to uncover what types of errors were common and what cues might have been used by these L2 subjects in restoring the test passages.

5. Information from the interviews was used to verify the results of the error analysis.

3.6 PILOT STUDY

3.6.1 PART I

The first part of the pilot study was conducted with a group of 23 L1 students (a mixed class of grades 11-12[4] in a composite high school in Edmonton, Alberta) in December, 1985. Each student was assigned to complete one of the four test forms. The tests were administered at random so that students sitting next to one another did not have the same form. Each form was taken by roughly the same number of students. The purpose of the study was to examine the appropriateness of each of the six passages, in all four forms. Table 3.3 shows a summary of results in percentage.

The results in Table 3.3 do not support Klein-Braley & Raatz's assumption that adult native speakers should achieve a perfect score. The scores ranged from 85.2% (MC-Test 2) to 97.9% (C-Test 1), with a fair amount of variability in MC-Test 2 (*S.D.*=16.6). These results were not so surprising, however, for it had been proposed from the beginning that adult native speakers did not necessarily achieve perfect scores.[5] On the basis of this finding, it was expected that the L1 sample of the main study would achieve an average score of at least 80% for all four forms.

3.6.2 PART II

The second part of the pilot study was carried out with a group of 13 L2 students in an adult education center in Edmonton. These students were classified as level 3+/4- (or intermediate level)[6] of that center. Each subject was randomly administered one of the four test forms, following the same procedure as that used for the L1 subjects. On the following

---

[4] Based on German L1 group data, Klein-Braley & Raatz (1984) claim that after the tenth grade (or age 16), "the L1 learner turns into the adult educated native speaker and achieves a perfect score on all tests" (p. 140).

[5] Under the assumption that native speakers differ in their reading ability, the scores of native speakers were used only as a general index of difficulty.

[6] ESL students in this center are classified into 7 levels of proficiency: beginning level, levels 1 to 5 and TOEFL preparation level.

Table 3.3

Mean Percent Scores and S.D. of Tests (L1 Data: Pilot Study)

Form	n	M	SD
C-Test 1	6	97.9	2.1
MC-Test 1	6	89.9	9.7
C-Test 2	6	94.1	5.1
MC-Test 2	5	85.2	16.6

day, the same group of subjects was administered the Michigan Test. These two tasks were completed during the second week of January, 1986, which was also the second week of the first half of winter session at the center.

The objectives of the second part of the pilot study were:

1. to examine the appropriateness of the test passages to be used with the sample of the main study;
2. to determine whether or not there should be a time limit for the subjects to complete the C-Test/MC-Test;
3. to investigate proficiency of the L2 subjects on the basis of their performance on the Michigan Test;
4. to examine a trend regarding how well the C-Test and MC-Test scores correlate with the Michigan Test scores; and
5. to try out the "information sheet" in order to determine whether there is need to add/omit or combine/revise any of the items.

The data obtained from the second part of the pilot study answered many questions.

1. As shown in Table 3.4, the mean percent scores of the tests range from 60.4% to 81.8%, suggesting that the tests were too easy for the subjects in this sample.[7] Since it is recommended that a minimum of 100 items per test is sufficient, two subtests were omitted. It seemed that the first two subtests should have been discarded so that the tests would be more difficult. However, when another aspect, the subject matter of the passages, was taken into consideration, it was decided that subtests 1 and 5 be omitted.

2. It was observed during the pilot study that even when there was no time limit for the subjects to complete the tests, some subjects did not finish many of the items. The tests were therefore be administered as "power" tests (i.e., having no time limit) rather than "speed" tests.

---

[7] Klein-Braley & Raatz (1984) suggest that tests should normally produce a mean difficulty level of 50 % for the target group envisaged; however, they also argue that tests which are far too difficult or far too easy for the target group do still produce acceptable reliability and validity coefficients (p. 140).

Table 3.4

Mean Percent Scores and S.D. of Tests (L2 Data: Pilot Study)

Form	n	M	SD
C-Test 1	3	81.8	6.5
MC-Test 1	3	60.4	11.7
C-Test 2	3	71.1	19.0
MC-Test 2	4	65.7	9.7

Table 3.5

Comparison between the C-Test/MC-Test Scores and the Michigan Test Scores of L2 Subjects

Form		C/MC Score	Michigan Score
C-Test 1		89.3	52.0
		78.0	44.0
		78.0	42.0
	M	81.8	46.0
	SD	6.5	5.3
MC-Test 1		70.0	49.0
		64.0	34.0
		47.3	38.0
	M	60.4	40.3
	SD	11.8	7.8
C-Test 2		90.0	59.0
		71.3	38.0
		52.0	36.0
	M	71.1	44.3
	SD	19.0	12.7
MC-Test 2		74.7	62.0
		69.3	51.1
		66.7	38.0
		52.0	47.0
	M	65.7	49.5
	SD	9.7	9.9
Total	M		45.4
	SD		8.9

3. Table 3.5 shows the Michigan Test scores (equated) of individual subjects in comparison with their C-Test/MC-Test scores. The Michigan Test scores range from 34 to 62, with a mean score of 45.4 and a *S.D.* of 8.9. The average for each group is 46 for C-Test 1, 40.3 for MC-Test 1, 44.3 for C-Test 2 and 49.5 for MC-Test 2. Since the mean scores of the whole group and each individual group fell a little below 50%, the subjects in this sample could probably be considered as a "lower intermediate" level.

4. Data in Table 3 also show that the C-Test/MC-Test correlated moderately well with the Michigan Test. Since the number of subjects who took each form of test was very small (three to four), correlation coefficients were not computed.

5. Although the information sheet already looked suitable in terms of format (all were fill-in-the-blank items), a few items were revised for a practical purpose before it was used for the main study. Most of the information remained the same.

Chapter 4

FINDINGS AND INTERPRETATION

In this chapter, results of the analyses of L1 and L2 data will be presented. The results are divided into four sections. Section one (4.1), *Comparability*, deals with item difficulty and item discrimination of the two test formats. In the second section (4.2), *Reliability*, the internal consistency of each form and each subtest of the four test forms, is described. Section three (4.3), *Validity*, is involved with factor analysis results from L1 data and correlational analysis results from L2 data. In the final section (4.4), *Error Analysis and Interviews*, the error patterns that emerge from the error analysis of L2 data and the reading strategies employed by those students, based on the interview data, are listed and discussed.

4.1 COMPARABILITY

The report of results in this section is divided into two parts. The first part deals with item difficulty and the second part discusses item discrimination.

4.1.1 ITEM DIFFICULTY

4.1.1.1 OVERALL RESULTS (L1 DATA)

The means and standard deviations of the four forms of test for L1 data are displayed in Table 4.1. The mean percent scores range from 82.80 to 93.93, confirming that the test forms are usable for the L2 sample.[1].

The MC-Test format appears to be relatively more difficult than the C-Test format. However, there is no significant difference in the mean scores between Form 1 and Form 2 of either format. This suggests that different starting points of mutilations do not affect the difficulty levels of the tests. The standard deviations indicate that the

---

[1] Klein-Braley & Raatz (1984) suggest that adult native speakers should obtain virtually perfect scores. However, based on the results of the pilot study by the author of this study, this L1 sample was expected to obtain at least an average of 80%.

Table 4.1

Mean Percent Scores and Standard Deviations of the C-Tests and MC-Tests (L1 Data)

Form	n	M	SD
C-Test 1	102	92.71	8.13
MC-Test 1	95	83.67	14.97
C-Test 2	102	93.98	7.03
MC-Test 2	90	82.80	13.83

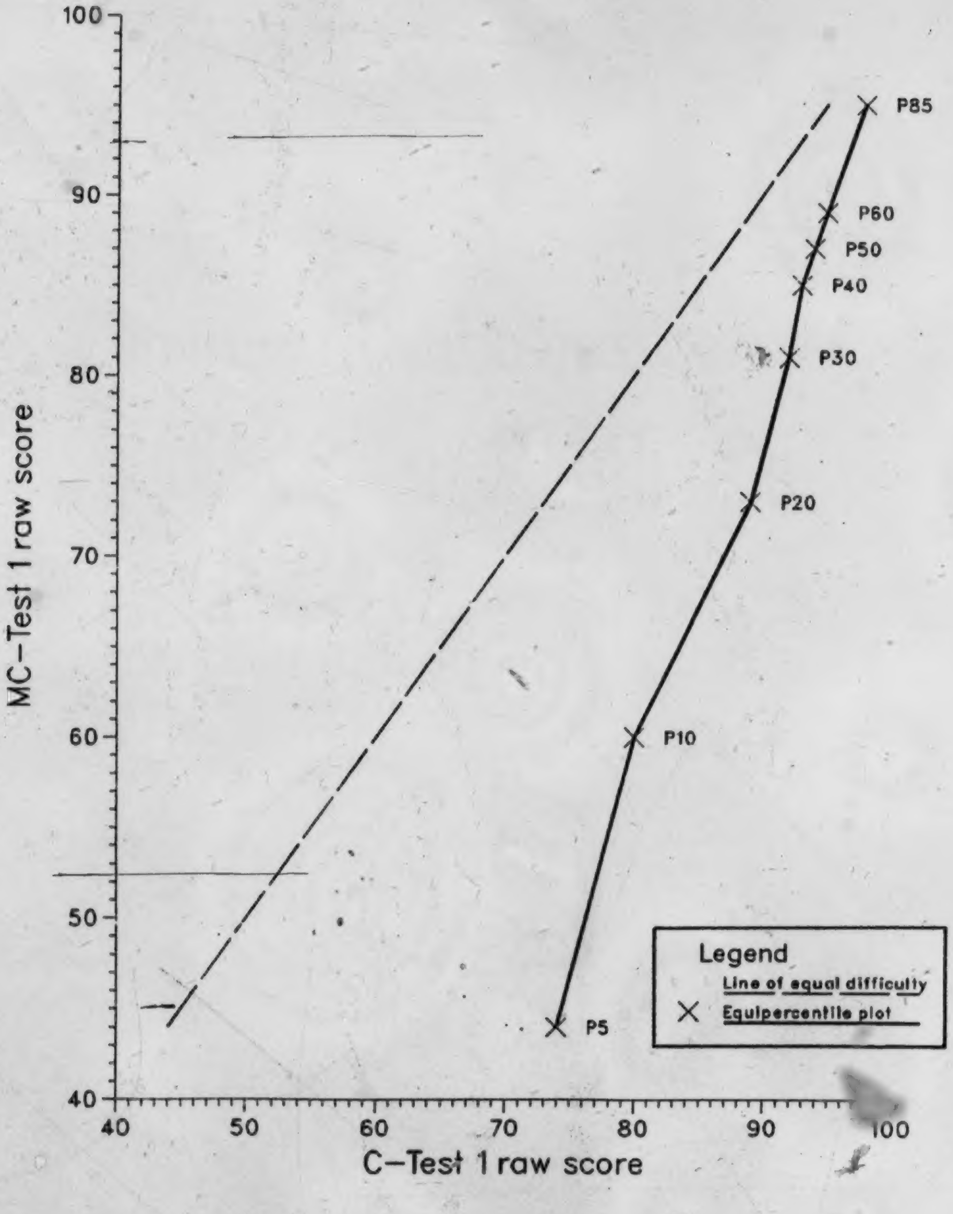


Figure 4.1 Equipercentile Plot of C-Test 1 and MC-Test 1: L1 Data

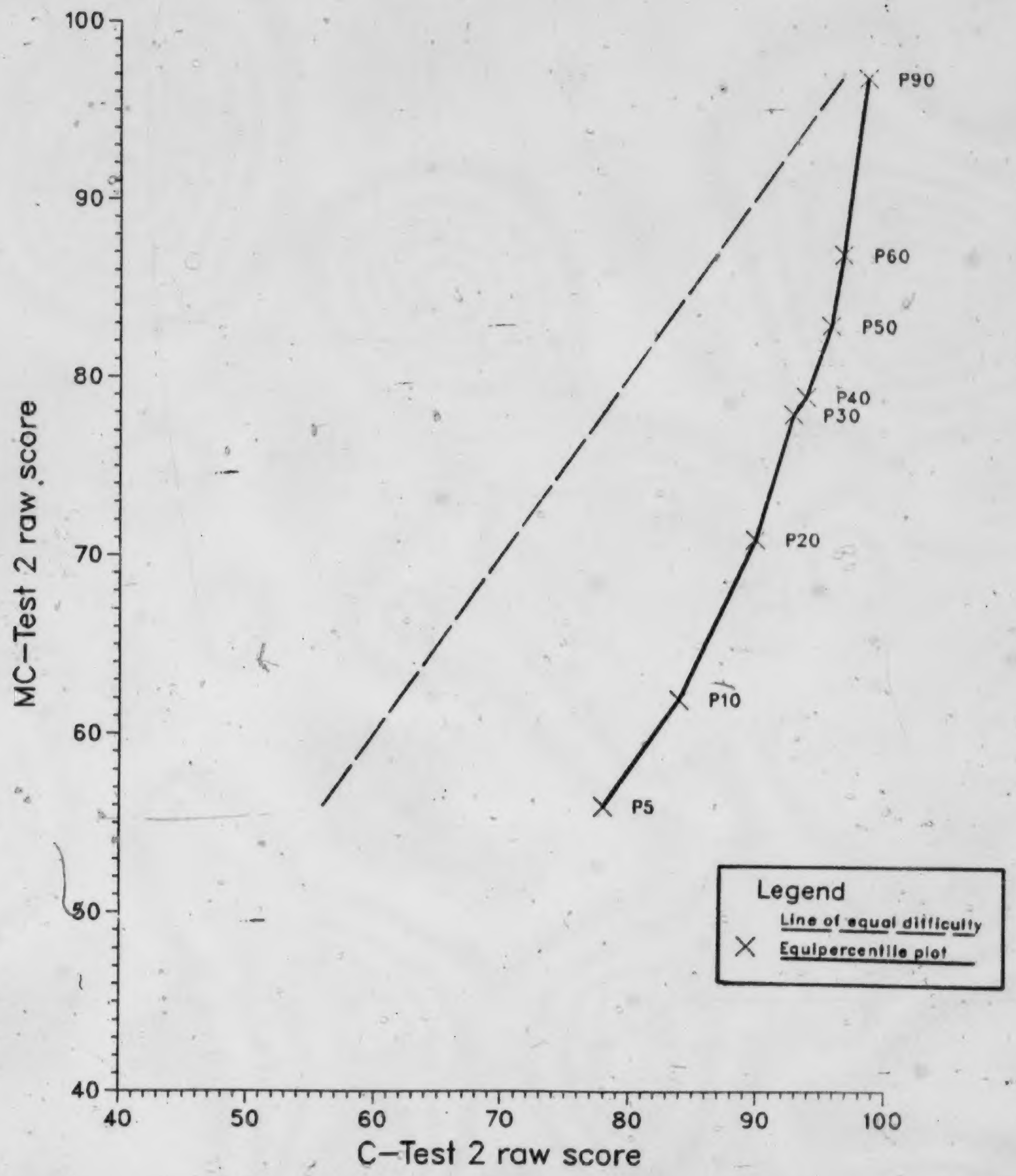


Figure 4.2 Equipercentile Plot of C-Test 2 and MC-Test 2: L1 Data

scores for the MC-Test format are more widely spread than those for the C-Test format, suggesting that the MC-Test format differentiates the student ability levels better than the C-Test format.

The results are also illustrated graphically by the equipercentile plots in Figures 4.1 and 4.2. At every percentile point, the raw score for the C-Test format is always higher than its corresponding raw score for the MC-Test format, which illustrates that at every ability level, the MC-Test format is more difficult than the C-Test format. In Figure 4.1, the shape of the plot, which almost forms a straight line and falls to the right, illustrates that the higher the percentile point, the less the difference between the raw scores for the two test formats. The shape of the plot in Figure 4.2, though also falling to the right, forms a curve instead of a straight line which demonstrates a similar trend as in Figure 4.1.

4.1.1.2 SUBTEST RESULTS (L1 DATA)

The patterns of the subtest results are consistent with the overall results. Table 4.2 reveals that in every single pair, the MC-Test subtest is always more difficult than its corresponding C-Test subtest. As well, the standard deviations of subtests indicate that the scores for the MC-Test subtests are more widely spread than those for their corresponding C-Test subtests. Furthermore, the results of the item analysis (see Tables B-1 through B-4, Appendix B) show that in most cases, the MC-Test items are more difficult than their corresponding C-Test items.

4.1.1.3 OVERALL RESULTS (L2 DATA)

Table 4.3 displays the means and standard deviations for L2 data. The results indicate that the four forms of tests were not too difficult for this L2 sample, as the mean percent scores range from 64.69 to 75.92.[9] For both test pairs, the standard deviation is greater for the MC-Test than for the C-Test. However, the difference in the

---

[9] Klein-Braley & Raatz (1984) suggest that C-Tests should be "revised until they produce a mean difficulty level of 50 per cent for the target group envisaged" (p. 140).

Table 4.2

Mean Scores and Standard Deviations for C-Test and MC-Test Subtests (L1 Data)

Form	n		Subtest			
			1	2	3	4
C-Test 1	102		C1S1	C1S2	C1S3	C1S4
		M	23.53	23.95	23.51	21.72
		SD	2.13	1.78	2.17	3.38
MC-Test 1	95		MC1S1	MC1S2	MC1S3	MC1S4
		M	21.49	23.22	20.84	18.12
		SD	4.74	2.59	4.71	5.65
C-Test 2	102		C2S1	C2S2	C2S3	C2S4
		M	24.11	23.47	23.60	22.80
		SD	1.44	2.17	1.67	3.24
MC-Test 2	90		MC2S1	MC2S2	MC2S3	MC2S4
		M	22.53	20.22	22.31	17.73
		SD	3.27	4.66	2.68	5.47

Note: Maximum possible score for each subtest = 25.

Table 4.3

Mean Percent Scores and Standard Deviations of the C-Tests and MC-Tests (L2 Data)

Form	n	M	SD
C-Test 1	26	75.08	10.25
MC-Test 1	26	64.69	17.99
C-Test 2	26	75.92	16.70
MC-Test 2	26	66.85	17.66

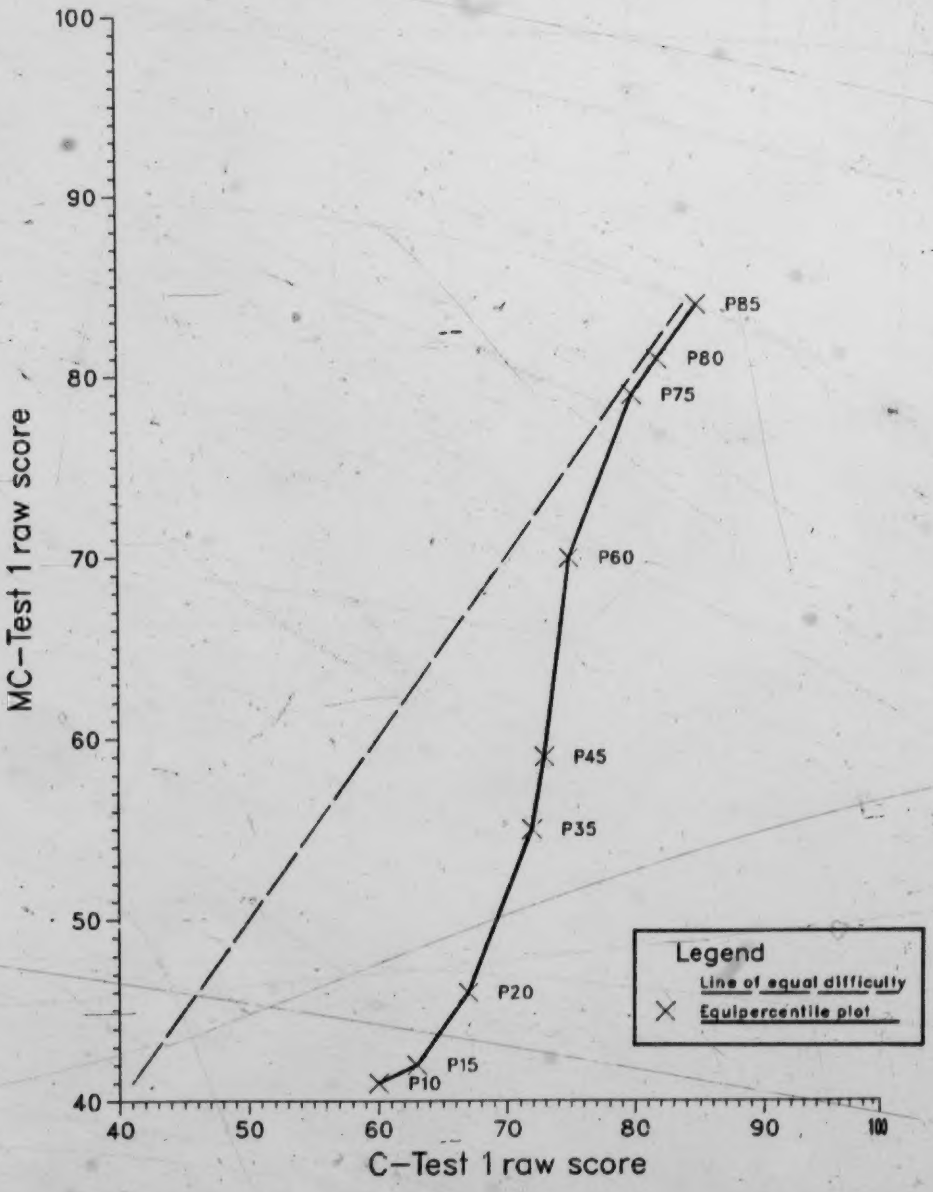


Figure 4.3 Equipercentile Plot of C-Test 1 and MC-Test 1: L2 Data

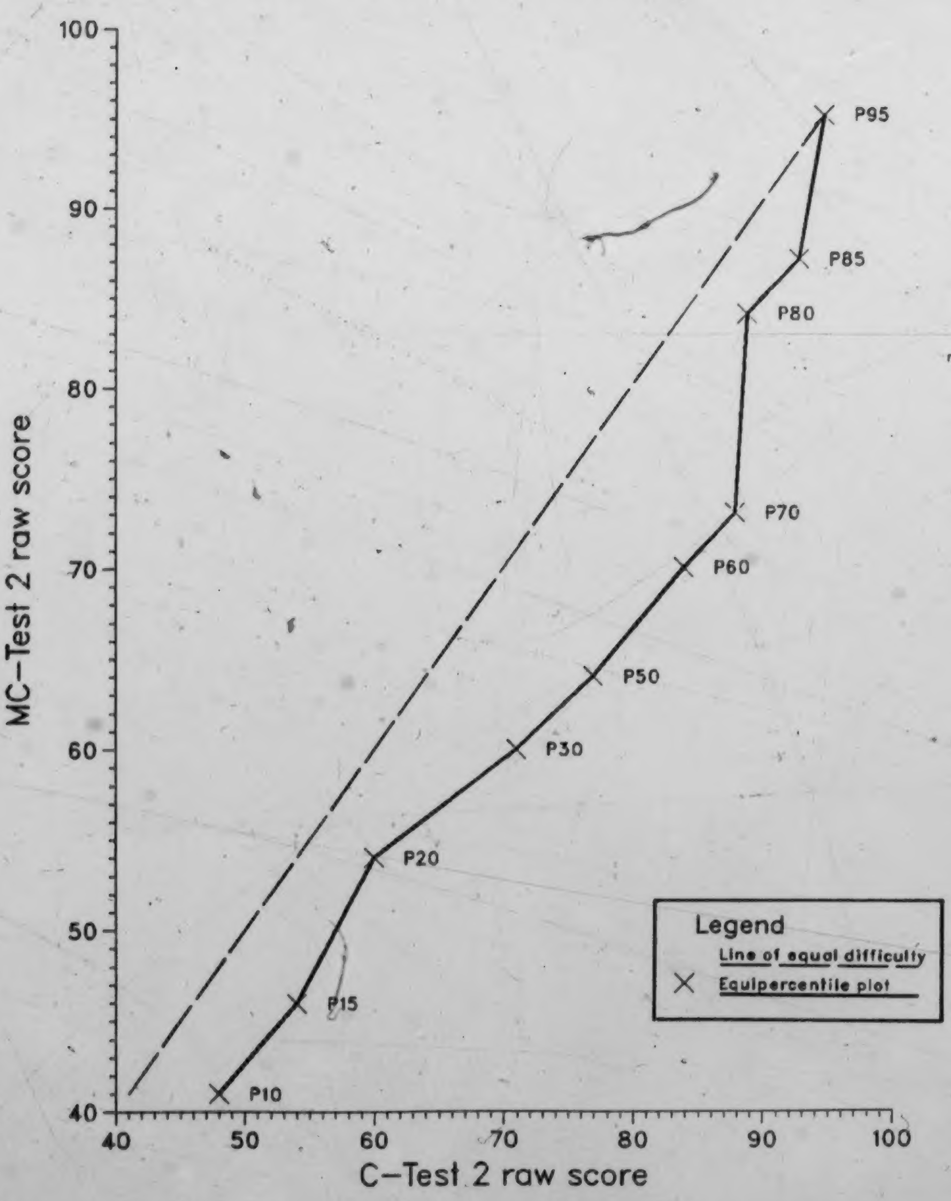


Figure 4.4 Equipercentile Plot of C-Test 2 and MC-Test 2: L2 Data

variability of individual scores, though consistent with L1 data, is much greater within the C-Test 1/MC-Test 1 pair than within the C-Test 2/MC-Test 2 pair. Therefore, the MC-Test differentiates the student ability levels better than the C-Test format. In general, the results are consistent with those for L1 data, because the MC-Test format appears to be more difficult than the C-Test format and the difference between the mean scores of Form 1 and Form 2 is insignificant.

The graphic presentations shown in Figures 4.3 and 4.4, though having somewhat different shapes from those for L1 data, reflect a similar trend. At every ability level (except at P95 in Figure 4.4), the MC-Test format appears to be more difficult than the C-Test format. The equipercentile plot in Figure 4.3 forms a curve similar to that in Figure 4.2 for L1 data, but from P75 to P85 the plot constitutes a straight line parallel to the line of equal difficulty. This suggests that the differences between the raw scores for C-Test 1 and MC-Test 1 are generally greater at lower percentile points than at higher percentile points. From P75 to P85 the differences appear to be minimal and constant.

In Figure 4.3, the plot forms a wavy line which demonstrates a trend toward parallelism with the line of equal difficulty. This shape illustrates that the differences between the raw scores for C-Test 2 and MC-Test 2 are relatively stable, except at the highest percentile point (P95) where there is no difference. The plot thus indicates that at all other ability levels, MC-Test 2 is more difficult than C-Test 2; only at the highest ability level (P95), the two forms of tests appear to be of equal difficulty.

4.1.1.4 SUBTEST RESULTS (L2 DATA)

The data for subtests in Table 4.4 show that of all the eight pairs for L2 data, there is only one pair (C2S1/MC2S1) where the MC-Test subtest is less difficult than its corresponding C-Test subtest. As well, MC2S1 is the only MC-Test subtest whose standard deviation is lower than its C-Test counterpart. The data thus suggest that in general the individual subtest scores for L2 data are more widely spread for the MC-Test format than for the C-Test format. That most MC-Test items are more difficult than their C-Test counterparts is supported by the results of the item analysis in Tables B-5 to

Table 4.4

Mean Scores and Standard Deviations for C-Test and MC-Test Subtests (L2 Data)

Form	n		Subtest			
			1	2	3	4
C-Test 1	26		C1S1	C1S2	C1S3	C1S4
		M	20.96	20.27	18.12	15.73
		SD	2.88	2.84	3.50	3.54
MC-Test 1	26		MC1S1	MC1S2	MC1S3	MC1S4
		M	19.62	17.35	15.23	12.50
		SD	4.54	4.96	5.78	5.40
C-Test 2	26		C2S1	C2S2	C2S3	C2S4
		M	21.12	18.23	19.96	16.62
		SD	3.50	4.46	4.33	6.27
MC-Test 2	26		MC2S1	MC2S2	MC2S3	MC2S4
		M	21.23	14.77	18.23	12.62
		SD	3.00	5.32	5.79	7.19

Note: Maximum possible score for each subtest = 25.

B-8, Appendix B.

4.1.2 SUMMARY OF ITEM DIFFICULTY

The item difficulty statistics demonstrate that test items are normally easier when the second half of word is deleted (C-Test format) than when the first half of the word is deleted (MC-Test format). In other words, the information given by the first part of words appears to be more helpful to the respondents than that given by the second part. On the other hand, no significant difference was found between Form 1 and Form 2 (C-Test 1 vs. C-Test 2; MC-Test 1 vs. MC-Test 2), suggesting that different starting points of mutilations do not affect the difficulty levels of the tests. To account for the results, the structure of English words should be seriously taken into consideration.

As discussed in Chapter 1, the structure of content words in English is such that the core or the semantic substance is usually contained in the front part of the word. This observation is supported by the fact that the vast majority of content words can have a structure of *stem and suffix(es)*. Thus, for the test items constructed from words having this type of structure, the semantic information (primarily in the first part of the words) must be more salient than the syntactic/structural information (which may be present in the second part).

However, the test items have been constructed from not only content words but function words as well. In fact, the proportions of content words in the passages to function words are 43:57 for the C-Test 1/MC-Test 1 items and 35:65 for the C-Test 2/MC-Test 2 items. It is therefore imperative to consider function words and their structure as well.

4.1.2.1 STRUCTURE OF FUNCTION WORDS

Function words are normally single morphemes, having only one or two syllables. According to *The American Heritage Word Frequency Book* (Carroll, et al., 1971), of the top 100 most frequent words, 84 are function words and only eight of these function words have more than one syllable. Since they are mostly "little" words of one or two syllables, the majority of them are composed of less than six letters. To correspond with the guidelines for deletion (Chapter 1, Section 1.7), function words are classified into four types of orthographic structures.[10] The four types along with examples[11] in the order of their relative frequencies are listed in Table 4.5.

As shown in Table 4.5, there are only two single-letter words in English (a & I). Because for a single-letter-word item, the whole word is deleted, the word a (or I) should be of equal difficulty in both test formats.

For words of more than one letter, the difficulty levels of the test items for the two test formats should mainly be due to the orthographic constraint of the words. For example, the C-Test item for the word we (w-) should be easier than its corresponding MC-Test item (-e), because there are more possible responses for the MC-Test item (e.g., we, he, me, be). Another important aspect that may account for the difference in difficulty levels of the items for the two test formats is the saliency of the undeleted parts of the items. Since English is a left-to-right language, it is reasonable to expect that the first part of the word is more recognizable than the second part. These observations correspond with the conclusions regarding the important parts of words in cueing word recognition (Weaver, 1980) which indicate that:

> 1. Consonants are more important than vowels.
> 2. Beginnings of words are more important than middles and ends, and ends are more important than middles. ... (p. 50)

A study on spelling errors also indicates that "relatively few spelling errors on words within the reading vocabulary involve the first two or three letters -- letters that are almost always a part of the word-recognition information that has been stored" (Kooi, Schutz, & Baker, 1965; cited in Simon & Simon, 1973, p. 130).

---

[10] Stanovich (1980) points out: "Intraword redundancy arises because of the sequential and position-specific constraints on the letters within words. Thus, written language is orthographically structured" (p. 37).

[11] The examples are taken from the "Rank List" of *The American Heritage Word Frequency Book* (Carroll, et al., 1971).

Table 4.5

Orthographic Structures of Function Words

Orthographic Structure	Examples
Single letter	a, I.
Two or three letters	the, of, and, to, in, is, you, it, he, for, was, on , are, as, his, at, be, or, by, one, had, not, but, all, we, can, an, if, do, how, up, out, she, so, has, her, two, him, no, its, who, now, my, did, may, etc.
Four or five letters	that, with, they, this, from, have, what, were, when, there, your, which, their, will, each, about, them, then, many, some, these, would, other, into, more, could, than, first, been, over, down, only, very, after, just, where, most, etc.
Six or more letters	through, before, around, another, because, between, should, something, without, during, however, several, etc.

4.1.2.2 ANOMALOUS ITEMS

Despite the fact that in general the C-Test items are easier than the MC-Test items (see Tables B-1 to B-8, Appendix B), there are certain items where the opposite is true. These anomalous items are listed in Tables 4.6 and 4.7.

Table 4.6 reveals that of the 100 items, there are 10 items for L1 data and 21 items for L2 data where the C-Test 1 items appear to be more difficult than their corresponding MC-Test items. The results given in Table 4.7 indicate that of the 100 items listed, there are 4 such anomalies for L1 data and 17 for L2 data in the C-Test 2/MC-Test 2 pair. It is important to note that the majority of these anomalous items are function words. Of the 41 anomalous items from both sets of data[11], only nine items (or 22.0%) are content words.

A close examination of the data reflects that most of these words appear to have different structures from those previously discussed. The responses for the MC-Test items of these words tend to be more constrained by orthographic information than those for their corresponding C-Test items. A characteristic example for this is the words up and or. When the last part is given (item #10, MC1S4 & item #14, MC2S3), the responses are constrained only to the words up and or respectively, because there are no other two-letter English words ending in p and r. On the other hand, when the first part is given (item #10, C1S4 & item #14, C2S3), more than one word are possible (i.e., up & us; or, of, on, oh, & ox).

Another obvious example is the four- or five-letter words beginning with th. The responses for the word this, for example, are much more open in the C-Test format than in the MC-Test format. When the second part is given (items #2 & #19, MC1S4), the possiblity for the responses is limited to only the word this, whereas their corresponding C-Test items allow for many possible words (e.g., this, that, than, then, they, them, & thus).

---

[11] Duplicate items from L1 and L2 data are counted once.

Table 4.6

Items for Which the Difficulty Indices in C-Test 1 are Smaller than in MC-Test 1

Subtest	Item	
	L1 Data	L2 Data
S1	5. her	5. her
	10. them	10. them
	-	23. said
S2	-	2. is
	-	9. It
	-	11. room
	-	15. that
	17. any	17. any
	21. that	21. that
	-	22. has
S 3	-	3. last
	-	9. only
	-	11. in
	-	12. lungs
	18. difficult	-
	-	25. only
S4	2. this	2. this
	-	3. that
	5. lived	-
	10. up	10. up
	-	16. in
	19. this	19. this
	20. their	20. their
Total	10	21

Table 4.7

Items for Which the Difficulty Indices in C-Test 2 are Smaller than in MC-Test 2

Subtest	Item	
	L1 Data	L2 Data
S1	6. first	6. first
	-	9. of
	-	10. was
	-	17. will
	-	20. for
	-	22. then
	-	23. Yes
S2	-	8. wall
	-	12. that
	19. than	19. than
	-	25. that
S3	-	10. stored
	-	11. the
	-	14. or
	22. As	-
S4	-	2. is
	-	7. is
	9. grew	9. grew
Total	10	21

It is thus reasonable to conclude that the orthographic constraint, the saliency of each part of words, together with the context of the text (e.g., word order) have all contributed to the results of the tests. For most content words the semantic saliency seems to play a more important role than the syntactic/structural saliency. On the other hand, for most function words the orthographic constraint and the saliency of each part of the words are significant components for word identification. The context of the text, especially the word order cue, should play an equal role in facilitating the responses for both content-word and function-word items.

4.1.3 ITEM DISCRIMINATION

In order to determine whether the structures of the C-Test and its corresponding MC-Test are similar in terms of discrimination power, items in each pair of subtests have been cross-tabulated within four intervals of discrimination power indices. The results are displayed in Tables C-1 through C-16 (Appendix C).

4.1.3.1 L1 DATA

The results of cross-tabulations in Tables C-1 to C-8 have been summarized in Tables 4.8 and 4.9. The findings in Table 4.8 reveal that 33 of the 100 items in the C-Test 1/MC-Test 1 pair share the same intervals of discrimination indices for both forms. Of the 100 items, 67 items in C-Test 1 and 82 items in MC-Test 1 have discrimination indices greater than .50. Subtest 4 appears to have the largest percentage of common items (13 items or 52%). Subtest 2 is the only subtest in this pair where there are more items with high discrimination power ($r > .50$) in C-Test 1 than in MC-Test 1.

The data indicate that only 33% of the items from both forms fall into the same discrimination intervals. This suggests that generally C-Test items and MC-Test items do not relate to the total test scores in the same way. Thus, it supports the results of the item difficulty that these two types of tests are functionally different. They are not tapping the same abilities. MC-Test 1 appears to be superior to C-Test 1, since there are

Table 4.8

Summary of Item Discrimination Statistics for C-Test 1 and MC-Test 1 (L1 Data)

Subtest	Common Items	Items with r >.50	
		C1	MC1
S1	7	13	23
S2	6	16	13
S3	7	17	22
S4	13	21	24
Total	33	67	82

Table 4.9

Summary of Item Discrimination Statistics for C-Test 2 and MC-Test 2 (L1 Data)

Subtest	Common Items	Items with r >.50	
		C2	MC2
S1	10	8	16
S2	7	13	23
S3	9	11	14
S4	12	18	21
Total	38	50	74

more items with high discrimination power in the former than in the latter.

Table 4.9 indicates that 38 of the 100 items in this pair fall into the same intervals of discrimination power for both forms. There are 50 items in C-Test 2 and 74 items in MC-Test 2 that have discrimination indices greater than .50. Subtest 4 in this pair has the largest number of common items (12 items or 48%). In all four subtests, the number of items with high discrimination power ($r > .50$) in MC-Test 2 is usually larger than that of its corresponding C-Test 2.

The results suggest that in this pair, 38% of the items seem to relate to the total test scores in the same way.This comfirms the conclusion that the two item types are functionally different. MC-Test 2 appears to be superior to C-Test 2 with regard to discrimination power.

4.1.3.2 L2 DATA

The results for L2 data (Tables C-9 to C-16) have been summarized in Tables 4.10 and 4.11.

As shown in Table 4.10, for L2 data there are only 23 common items in this pair. Of the 100 items, 36 items in C-Test 1 and 63 items in MC-Test 1 have high discrimination power ($r > .50$). An interesting consideration is that, unlike the preceding reports on L1 data, Subtest 4 in this set has the smallest percentage of common items (only 3 items or 12%). This is clearly because there are 20 of the MC1S4 items but only 7 of the C1S4 items which have high discrimination power.

The results suggest that for L2 data, MC-Test 1 is far superior to C-Test 1 with respect to item discrimination power. Again, in this pair only 23% of the items appear to relate to the total test scores in the same way, suggesting that they are not tapping the same language abilities.

The data in Table 4.11 show that 50 items in the C-Test 2/MC-Test 2 pair share the same intervals of discrimination indices. Of the 100 items, 71 items in C-Test 2 and 64 in MC-Test 2 have discrimination indices greater than .50. Subtest 4 in this pair, has the largest number of common items (18 items or 72%). Unlike L1 data, there are more

Table 4.10

Summary of Item Discrimination Statistics for C-Test 1 and MC-Test 1 (L2 Data)

Subtest	Common Items	Items with r >.50	
		C1	MC1
S1	8	9	11
S2	7	11	16
S3	5	9	16
S4	3	7	20
Total	23	36	63

Table 4.11

Summary of Item Discrimination Statistics for C-Test 2 and MC-Test 2 (L2 Data)

Subtest	Common Items	Items with r >.50	
		C2	MC2
S1	11	12	5
S2	9	18	18
S3	12	21	19
S4	18	20	22
Total	50	71	64

items with high discrimination power ($r > .50$) in C-Test 2 than in MC-Test 2 for L2 data. However, Subtest 4 is the only subtest in which the number of items with high discrimination power in MC-Test 2 surpasses the number of those in C-Test 2.

The results demonstrate that 50% of the items in this pair appear to relate to the total test scores in the same way. This implies that, in general, C-Test items and MC-Test items are functionally different. For this set of data, however, C-Test 2 appears to be superior to MC-Test 2 in reference to item discrimination power.

4.1.4 SUMMARY OF ITEM DISCRIMINATION

The information obtained from the item discrimination statistics indicates that the C-Test format and MC-Test format are remarkably different in their structures of item discrimination. For L1 data, 33% of the items in the C-Test 1/MC-Test 1 pair and 38% in the C-Test 2/MC-Test 2 pair appear to relate to the total test scores in the same way. However, for L2 data only 23% of the items in the C-Test 1/MC-Test 1 pair but 50% in the C-Test 2/MC-Test 2 pair are found to relate to the total test scores in a similar way.

The difference between the two test pairs is that in the C-Test 2/MC-Test 2 pair the deletion begins at the third word, instead of the second word of the second sentence as in the C-Test 1/MC-Test 1 pair. The results thus suggest that the different starting points of deletion have more effect on the item discrimination structures in L2 data than in L1 data. Whereas the difference between the common items of the two test pairs is only 5% for L1 data, it is 27% for L2 data. This is partly due to the fact that the variability of the scores is greater in L2 than in L1 data.

Findings indicate that MC-Test 1 is superior to C-Test 1 with regard to item discrimination power. There are more items with high discrimination power in MC-Test 1 than in C-Test 1 for both L1 and L2 data. The results obtained for C-Test 2 and MC-Test 2, on the other hand, are not consistent for the two sets of data. For L1 data, there are more items with high discrimination power in MC-Test 2 than in C-Test 2, whereas for L2 data, the opposite is true. MC-Test 2 is thus superior to C-Test 2 in reference to item discrimination power for the L1 sample but not for the L2 sample.

In summary, it is apparent that the two types of items are functionally different, since they do not appear to relate to the total test scores in the same way. This suggests that they are not tapping the same abilities. It should be remembered that the difference between the C-Test and the MC-Test is that of the word part deleted. Whereas the second part of a word is deleted in the C-Test, it is the first part that is deleted in the MC-Test. As mentioned in the discussion of the results of item difficulty, this makes a significant difference. When the second half is deleted (C-Test), generally there will be more semantic information present in the item than when the first half is deleted (MC-Test). On the other hand, MC-Test items will essentially contain more syntactic/structural information than C-Test items.The findings from item discrimination in this section together with the results of item difficulty discussed above firmly suggest that these two item types are different.

4.2 RELIABILITY

Information regarding the reliability of the four forms of tests (combined subtests) as well as each of the four subtests was obtained by using the Hoyt estimate of reliability. Cronbach's alpha coefficients for the four forms are also available.

4.2.1 L1 DATA

A summary of reliability statistics for the four forms of tests is displayed in Table 4.12. The coefficients of the Hoyt estimate of reliability are all greater than .90. The range of the standard errors of measurement is from 2.12 to 3.05. Both forms of the MC-Test format have larger coefficients than the corresponding C-Test forms. Cronbach's alpha coefficients range from .79 to .85. In reference to Cronbach's alpha coefficients, MC-Test 2 has a larger coefficient than C-Test 2 does. MC-Test 1, however, has a slightly smaller coefficient than C-Test 1. Cronbach's alpha is different from the Hoyt's reliability in that it has been calculated using the scores of the four subtests instead of the scores of individual items as for the Hoyt's reliability. Thus, Cronbach's alpha is an indication of how

Table 4.12

Summary of Reliability Statistics of C-Tests and MC-Tests (L1 Data)

Form	n	Hoyt	SE	Alpha
C-Test 1	102	0.93	2.17	0.85
MC-Test 1	95	0.96	2.94	0.84
C-Test 2	102	0.91	2.12	0.79
MC-Test 2	90	0.95	3.05	0.85

Note: Maximum possible score for each test = 100.

Table 4.13

Summary of Reliability Statistics for C-Test and MC-Test Subtests (L1 Data)

Form	n		Subtest			
			1	2	3	4
C-Test 1	102		C1S1	C1S2	C1S3	C1S4
		Hoyt	0.78	0.75	0.78	0.85
		SE	0.98	0.87	1.01	1.30
MC-Test 1	95		MC1S1	MC1S2	MC1S3	MC1S4
		Hoyt	0.92	0.82	0.90	0.92
		SE	1.35	1.07	1.45	1.57
C-Test 2	102		C2S1	C2S2	C2S3	C2S4
		Hoyt	0.67	0.75	0.63	0.86
		SE	0.82	1.06	0.99	1.20
MC-Test 2	90		MC2S1	MC2S2	MC2S3	MC2S4
		Hoyt	0.84	0.89	0.76	0.90
		SE	1.27	1.50	1.27	1.72

Note: Maximum possible score for each subtest = 25.

homogeneous the subtests are, whereas the Hoyt's reliability tells about the homogeneity of the items.

The reliability statistics for the subtests are displayed in Table 4.13. The results show that the coefficients of the Hoyt estimate of reliability for the C-Test subtests range from .63 to .86 (*S.E.*, 1.49 to 1.92), whereas for the MC-Test subtests the range is from .76 to .92 (*S.E.*, 1.59 to 1.91). In every single pair, the coefficient for the MC-Test subtest is always larger than that for the corresponding C-Test subtest. This is partly due to the fact that the item variances are greater for the MC-Test than the C-Test.

The findings suggest that for the L1 data, MC-Test items are more homogeneous than C-Test items. With respect to the homogeneity of subtests, however, the MC-Test format is superior to the C-Test format only in the C-Test 2/MC-Test 2 pair.

4.2.2 L2 DATA

Reliability statistics for L2 data have been summarized in Tables 4.14 and 4.15. The results in Table 4.14 indicate that the coefficients of the Hoyt estimate of reliability of the three forms are extremely high (coefficients=.96 for all three). C-Test 1 has a coefficient of .88, which can be considered high. The range of the standard errors of measurement is from 3.46 to 3.76. Cronbach's alpha coefficients range from .81 to .90. With regard to Cronbach's alpha coefficients, MC-Test 1 and C-Test 2 appear to be superior to C-Test 1 and MC-Test 2, respectively.

The results given in Table 4.15 show that the reliability coefficients for the C-Test subtests range from .65 to .91 (*S.E.*, 1.49 to 1.92), and for the MC-Test subtests, the range is from .71 to .94 (*S.E.*, 1.59 to 1.91). Of the eight pairs, C2S2/MC2S2 is the only pair where the coefficient of the C-Test format is greater than that of the MC-Test format.

Again, the findings from the L2 data reveal that MC-Test items are generally more homogeneous than C-Test items. In relation to the homogeneity of the subtests, on the other hand, only MC-Test 1 is superior to C-Test 1.

Table 4.14

Summary of Reliability Statistics of C-Tests and MC-Tests (L2 Data)

Form	n	Hoyt	SE	Alpha
C-Test 1	26	0.88	3.56	0.81
MC-Test 1	26	0.96	3.76	0.89
C-Test 2	26	0.96	3.46	0.90
MC-Test 2	26	0.96	3.65	0.81

Note: Maximum possible score for each test = 100.

Table 4.15

Summary of Reliability Statistics for C-Test and MC-Test Subtests (L2 Data)

Form	n		Subtest			
			1	2	3	4
C-Test 1	26		C1S1	C1S2	C1S3	C1S4
		Hoyt	0.70	0.65	0.72	0.69
		SE	1.54	1.64	1.83	1.92
MC-Test 1	26		MC1S1	MC1S2	MC1S3	MC1S4
		Hoyt	0.85	0.86	0.89	0.89
		SE	1.73	1.83	1.91	1.76
C-Test 2	26		C2S1	C2S2	C2S3	C2S4
		Hoyt	0.81	0.84	0.86	0.91
		SE	1.49	1.73	1.61	1.83
MC-Test 2	26		MC2S1	MC2S2	MC2S3	MC2S4
		Hoyt	0.71	0.89	0.91	0.94
		SE	1.59	1.74	1.67	1.78

Note: Maximum possible score for each subtest = 25.

4.2.3 SUMMARY OF RELIABILITY

The results from both L1 and L2 data demonstrate that all four forms of the tests are satisfactorily reliable.[13] The high coefficients of the Hoyt estimate of reliability suggest that the degree to which each test item taps the psychological trait the test is measuring is high for all four forms; that is, the test items are fairly homogeneous.

The reliability coefficients of the subtests indicate that generally the test items of the MC-Test subtests are more homogeneous than those of the C-Test subtests. In both L1 and L2 data, there is only one pair (C2S1/MC2S1, L2 data) for which the coefficient for the MC-Test subtest is smaller than for the C-Test subtest. The item discrimination statistics discussed in the previous section strongly support this observation, because there are almost always more items with high discrimination power in the MC-Test subtests than in the C-Test subtests. C2S1/MC2S1, L2 data, is the only pair where the MC-Test subtest has far fewer items with high discrimination power than the C-Test subtest.

Cronbach's alpha is an index of the consistency of the subtests; it indicates the homogeneity of the subtests (Nelson, 1974, pp. 279-280). The Cronbach's alpha coefficients in this study, which can be considered sufficiently high,[14] suggest that the degree to which the subtests are homogeneous is adequate for all the four forms. It is interesting to note that the results from the two sets of data are contradictory in that the alpha coefficient of MC-Test 1 is higher for L1 data but lower for L2 data than that of C-Test 1, whereas the MC-Test 2 coefficient is lower for L1 data but higher for L2 data than the C-Test 2 counterpart. This could be due to sampling variation.

Klein-Braley and Raatz (1984) used Cronbach's alpha to estimate the reliability of C-Tests. They viewed each subtest in the test as a "superitem" and concluded that "a

---

[13] Nelson (1974, p. 261) has suggested that whether or not the magnitude of a reliability coefficient is "good" depends largely upon the purpose of the test. He recommends that performance tests should have a reliability coefficient of .85 or higher.

[14] When subtest items correlate positively (within each subtest), the Hoyt value will exceed alpha. Normally, Hoyt will be expected to exceed alpha. (Nelson, 1974, p. 280). Klein-Braley & Raatz (1984) used Cronbach's alpha and considered each subtest as a "superitem". They have suggested a criterion of .80 or higher (p. 136).

[revised] C-Test consisting of only two superitems can provide reliable and valid information" (p. 140). However, they recommended that at least four texts be used in order to avoid text content bias.

The results of this study support Klein-Braley & Raatz's conclusion. For L2 data, all four forms, consisting of four subtests each, reached the .80 criterion. For L1 data, C-Test 2 ($r=.79$), although not reaching the criterion, is very close to it. The results are thus satisfactory. The contradiction of the results concerning the superiority of the two test formats in L1 and L2 data, however, seems to lead to a question of appropriateness in using Cronbach's alpha alone to estimate reliability of the C-Test/MC-Test.

4.3 VALIDITY

In this section the results will be presented in two parts. The first part (4.3.1) will describe the findings obtained from factor analyses of L1 data. In the second part (4.3.2), the correlations between the Michigan Test and the C-Test/MC-Test will be discussed.

4.3.1 FACTOR ANALYSIS (L1 DATA)

In order to discover whether the underlying structures for the C-Test and the corresponding MC-Test are the same, factor analysis was performed. Mulaik (1972) defines factor analysis as "a formal model about hypothetical component variables which account for the linear relationships that exist between variables" (p. 96). Factor analysis may be generally referred to as "techniques for anayzing test scores in terms of some number of underlying factors" (Hatch & Farhady, 1982, p. 255). The assumption on which factor analytic techniques are based is: "in any test there are probably one or more underlying traits [or factors] that are being assessed"; therefore, the purpose of factor analysis is "to construct these underlying factors and decompose these score variances in terms of correlation of the factors and the observed scores" (Hatch & Farhady, 1982, p. 255). Factor analysis will uncover how test items cluster and whether factor structures of the C-Test and its corresponding MC-Test are the same.

FACT20, a factor analysis computer program developed by DERS, The University of Alberta, was used to analyze the raw data from each subtest of the four test forms (i.e., 16 analyses). Principal components method and Varimax rotation[15] with a lower bound rotation of 2 factors were selected. The scree test (a plot of size of the eigenvalue of each factor) was conducted to determine the number of factors for each subtest. Table 4.16 shows the number of factors decided for each subtest.

4.3.1.1 ADJACENCY EFFECT

Because the results of the factor analysis reveal that the majority of items clustered into groups of adjacent items, a brief explanation of the term "Adjacency Effect" is necessary. Adjacency effect can be described as a testing phenomenon that occurs such that an item answered correctly (or incorrectly) will be followed by another item answered correctly (or incorrectly). An obvious example of adjacent effect is in answering a succesive set of reading comprehension questions where the answers to succeeding questions are based upon the correct answers of the preceding questions.

In C-Test and MC-Test formats of the cloze procedure, since half of every other word is deleted, it can be expected that many items will cluster into groups of adjacent items. This is more likely when the test items are constructed from an expression or phrase where a few certain words usually co-occur so that predicability of response is high. Examples of these are:

How old are you? (Items #11, #12 in C1S1 & MC1S1)

As a rule we . . . (Items #22, #23 in C2S3 & MC2S3)

Even though part of each underlined word is deleted, the respondents might answer an adjacent item correctly or incorrectly because of the common usage of the phrase. If the first item is answered correctly, the succeeding item is likely to be correct, but if the first item is answered incorrectly, it could affect the answer to the adjacent item.

---

[15] Mulaik (1972) suggested that Varimax rotation "is today perhaps the most used orthogonal-rotation procedure for attaining an approximation to simple structure"(p. 260).

Table 4.16

Number of Factors Extracted from Each Subtest of the C-Tests and MC-Tests

Form	Factors	Form	Factors
C-Test 1		MC-Test 1	
C1S1	5	MC1S1	5
C1S2	4	MC1S2	5
C1S3	5	MC1S3	4
C1S4	5	MC1S4	5
C-Test 2		MC-Test 2	
C2S1	4	MC2S1	6
C2S2	6	MC2S2	5
C2S3	5	MC2S3	6
C2S4	4	MC2S4	6

Adjacency effect also occurs when the answer to one item is a crucial key to the answer to the adjacent item. A good example is:

> " . . . I will try to think." She thought for . . . (Items #19, #20 in C1S1 &MCS1)

When a test respondent can figure out either the word think or thought, it is likely that he or she will easily get the correct answer for the other.

4.3.1.2 CONTENT WORDS VS. FUNCTION WORDS

The results of the factor analysis in this study also demonstrate that items frequently clustered into groups of words of the same categories. The two major categories which were identified include "Content Words" and "Function Words." The distinction between these two categories has been clearly made in Clark & Clark (1977):

> Content words are those that carry the principal meaning of a sentence. They name the objects, events, and characteristics that lie at the heart of the message the sentence is meant to convey. They include nouns, main verbs, adjectives, and most adverbs. Function words, in contrast, are those needed by the surface structure to glue the content words together, to indicate what goes with what and how. They include articles, pronouns, conjunctions, auxiliary verbs, and propositions. (p. 21).

The list of the different types of content and function words in Table 4.17 is as well from Clark & Clark (1977).

The distinction between content and function words is usually related to the concept of "openness" and "closedness" of the items.[16] Content words belong to the "open classes" because they are open for new members to be added, whereas function words, the "closed classes", are limited in number and new items are not normally created. The following illustration given by Quirk & Greenbaum (1973, p. 20) clearly demonstrates a contrast between the numbers of possible items for the "open classes" (columns *II* & *V*) and the "closed classes" (columns *I*, *III*, & *IV*).

---

[16] For detailed discussions, see Clark & Clark, 1977; Fromkin & Rodman, 1983; and especially Quirk & Greenbaum, 1973.

Table 4.17

The Major Word Classes in English

CONTENT WORDS

Nouns: dog, apple, matriarchy, elation, etc.
Verbs: go, receive, believe, trip, etc.
Adjectives: happy, naughty, pusillanimous, etc.
Adverbs: sadly, understandably, aptly, etc.

FUNCTION WORDS

Pronouns: I, you, she, . . . there.
Determiners: a, an, the, this, . . . some.
Quantifiers: much, a few, more, . . three.
Prepositions: in, on, beside, to, . . . of.
Intensifiers: very, too, a little, . . . quite.
Coordinate conjunctions: and, but, or, nor, also, so, yet.
Adverbial conjunctions: although, if, because . . . before.
Conjunctive adverbs: besides, nevertheless, . . . hence.
Relative pronouns: who, which, whose, . . . that.
Auxiliary verbs: can, may, have, . . . must.
Linking verbs: be.

Note: From *Psychology and language: An introduction to psycholinguistics* (p. 22) by H.H. Clark & E. V. Clark, 1977, New York: Harcourt Brace Jovanovich, Inc.

	I	*II*	*III*	*IV*	*V*
(John)	may	sit	by	this	fountain
	will	stare	at	that	tree
	must	read	from	___	window
	.	hurry	along		blackboard
	.	.	on		girl
	.	.	.		path
	___	.	.		.
			.		.
			___		.
					___

The concept of openness and closedness has been modified by Lee (1985) to explain the results of factor analyses of the cloze test. Lee took into account the context of the item which was supposed to belong to a closed class (or an open class) according to its part of speech but its loading was related to the open- (or closed-) type items. According to Lee, the word if, which is a function word, would be considered relatively open when the context allows for other possible words (e.g., as, when, because, since). On the other hand, the word opinion in the standard expression "in my opinion", which permits only limited responses, would be classified as closed. Since this extended concept of openness and closedness is psycholinguistically explicable, they will be adopted to account for the results of factor loadings in this study.

4.3.2 FACTOR SOLUTIONS

In this section, the report on the results from the 16 factor analyses will be presented in the order of subtests, in pairs (C-Test/MC-Test), factor by factor.[17] In order to make a clear comparison of the two different types of items constructed from the same words, the tables have been designed such that the words for the items are in the center and on either side are the numbers indicating the factors on which the test items load in each format. As well, each table will be accompanied by the test passages of both formats for the ease of reference. Importantly, since factor numbers are assigned in the order of the magnitude of common variance, the same factor numbers in different subtests and/or in different formats do not necessarily represent the same psychological trait. Thus, each factor will be described separately. In addition, following the descriptions of factors in each subtest pair, a cross-tabulation of items which load on each factor in the pair will be presented to compare their factor structures.

---

[17] The criterion for a cutoff point for factor loadings to be interpreted seems to be arbitrary. For example, Youngman (1979) suggested that traditionally, "loadings of over .50 are taken to define a factor while those over .30 can be used to add

Since the results are uniformly similar for all subtests, reptitiveness seems impossible to avoid. In order to reduce repitition and to avoid a lengthy description, test items will be fully identified only when it is essential for clarification.

4.3.2.1 C-TEST 1 SUBTEST 1 (C1S1)

Factor One. Table 4.18 shows that there are eight items loading on Factor 1; these are and, one, them, How, are, don't, I, and thought. Four of these items are adjacent items (. . . one of them was, "How old are you?"). All of the items except #20 (thought) are function words; however, thought in this particular context (. . . I will try to think." She thought for a minute . . .) is relatively "closed" in terms of possible responses, since it is constrained by the adjacent item think. Therefore, Factor 1 appears to be related to adjacency effect and closed-type items.

Factor Two. There are five items loading on Factor 2. Four of these (#17, #18, #19, & #20) are adjacent items, and four of the five items are content words. Only #17 (I) is a function word. The items loading on Factor 2 appear to be mainly

---

[17](cont'd) detail" (p. 107). On the other hand, Willemsen (1974) concluded that: "It is customary in factor-analysis literature for a loading of .33 to be the minimum absolute value to be interpreted" (p. 151). Since some criterion for a loading cutoff point must be adopted, the author has chosen to interpret loadings of .40 or greater, a practice widely accepted by current researchers.

Table 4.18

Items Loading 0.4 or Greater on Each Factor: Subtest 1 (C1S1 vs. MC1S1)

C1S1					Item	MC1S1				
Factor						Factor				
1	2	3	4	5		1	2	3	4	5
			4		1. was		2			5
		3			2. new			3		
1					3. and				4	
					4. not		2			
		3			5. her		2			
					6. he		2			
					7. asked			3		
			4		8. questions		2	3		
1		3			9. one		2	3		
1		3	4		10. them		2	3		
1					11. How	1	2			
1					12. are	1	2			
				5	13. Well	1		3		
				5	14. answered			3		
1					15. don't					5
					16. doctor	1				
1	2				17. I	1				
	2				18. try	1				
	2				19. think	1				
1	2				20. thought	1				
					21. a	1				
					22. and	1			4	
	2				23. said	1		3		
					24. I	1				
		3			25. now	1				

Note: Items 6, 7, 16, 22, and 24 of C1S1 were not employed in the analysis.

C-TEST 1

Passage 1 (C1S1)

A woman was having some trouble with her heart, so she went to see the doctor. He w_ _ a n_ _ doctor, a_ _ did n_ _ know h_ _, so h_ first as_ _ _ some ques_ _ _ _ _, and o_ _ of th_ _ was, "H_ _ old a_ _ you?"

"We_ _," she answ_ _ _ _, "I do_ _ _ remember, doc_ _ _, but _ will t_ _ to th_ _ _." She tho_ _ _ _ for _ minute a_ _ then sa_ _, "Yes, _ remember n_ _, doctor! When I married, I was eighteen years old, and my husband was thirty.

MC-TEST 1

Passage 1 (MC1S1)

A woman was having some trouble with her heart, so she went to see the doctor. He _ _s a _ _w doctor, _ _d did _ _t know _ _r, so _e first _ _ _ed some _ _ _ _tions, and _ _e of _ _em was, "_ _w old _ _e you?"

"_ _ll," she _ _ _ _ered, "I _ _ _'t remember, _ _ _tor, but _ will _ _y to _ _ _nk." She _ _ _ _ght for _ minute _ _d then _ _id, "Yes, _ remember _ _w, doctor! When I married, I was eighteen years old, and my husband was thirty.

related to adjacency effect and content words.

Factor Three. The five items loading on Factor 3 are new, her, one, them, and now. Of these five items, there is one pair of adjacent items, #9 and #10 (. . . one of them . . .), and this also appears in Factor 1. The only item which is not a function word is new; however, it is one among the top high frequency words[11] which are mostly function words. In addition, due to the context of this item (He was a new doctor, and did not know her, . . .), the responses to it are highly constrained. Thus, the items loading on this factor appear to relate to closed-type items.

Factor Four. Factor 4 contains three items, namely, was, questions, and them. If context is taken into consideration, it can be seen that questions is the referent of them. In addition, the responses for questions are constrained by asked and the orthographic information available in the item (. . . he first asked some questions, and one of them was. . .). This factor, therefore, can be considered as related to closed-type items.

Factor Five. There are only two items loading on Factor 5. These items are #13 and #14 ("Well," she answered, "I don't remember, doctor, . . .). The items loading on Factor 5 appear to be related to adjacency effect.

4.3.2.2 MC-TEST 1 SUBTEST 1 (MC1S1)

Factor One. There are 13 items loading on Factor 1 (Table 4.18) which can be divided into two groups of adjacent items. In the first group, there are only three items How, are, and Well. These occur in the context: "How old are you?" "Well," she answered, ... The items in the second group are all in the same paragraph and follow each other in consecutive order. It can be concluded that Factor 1 is related to adjacency effect.

---

[11] The word new is ranked 108th in *The American Heritage Word Frequency Book* (Carroll, et al., 1971). Of the top 108 words, 89 are function words.

Factor Two. Factor 2 is loaded with nine items. These items cluster into two groups of adjacent items and one isolated item. One group is #4, #5, and #6 (. . . and did not know her so he first asked . . .) and the other group is #8, #9, #10, #11, and #12 (. . . some questions, and one of them was, "How old are you?"). The isolated item is #1 (was). The nine items are all contained in one single sentence. Item #8 (questions), though being the only content word, has a characteristic of closedness. Thus, Factor 2 appears to be related to adjacency effect and closed-type items.

Factor Three. There are eight items loading on Factor 3. The items in this factor also indicate adjacency effect. The first group of adjacent items is #7, #8, #9, and #10. (. . . so he first asked some questions, and one of them was, . . .). The second group is #13 and #14 ("Well," she answered, . . .). The remaining items are #2 (new) and #23 (said). Factor 3 appears to be related mainly to adjacency effect and mostly content words.

Factor Four. This factor is loaded with two items, #3 & #22, which are the two occurrences of the word and. Thus, Factor 4 is related to repeated items, suggesting that when the first item is answered correctly, the second one would normally be correct too.

Factor Five. There are two items loading on this factor. These items are #1 (was) and #15 (don't). The items loading on Factor 5 is, therefore, related to function words or closed-type items.

4.3.2.3 COMPARISON OF FACTOR STRUCTURES (C1S1 & MC1S1)

Cross-tabulation of the items which load on each of the factors in C1S1 and MC1S1 is displayed in Table 4.19. Examination of the table reveals that four items (#11, #12, #17, & #20) are common to Factor 1 of both test forms. Items #9, #10, #11, and #12, which load on Factor 1 of C1S1, also load on Factor 2 of MC1S1. The five items (#17, #18, #19, #20,and #23) loading on Factor 2 of C1S1 are found to

Table 4.19

Cross-tabulation of Items Loading on Each Factor of C1S1 and MC1S1

MC1S1

	Factor	1	2	3	4	5	None
	1	11,12,17 20	9,10,11 12	9,10	3	15	
	2	17,18,19 20,23		23			
C1S1	3	25	5,9,10	2,9,10			
	4		1,8,10	8,10			
	5	13		13,14			
	None	16,21,22 24	4,6	7	22		

Note:

1) Number of items employed in factor analysis: C1S1 = 20; MC1S1 = 25.
2) Numerals in the cells are test item numbers.

load on Factor 1 of MC1S1. The factor-pairs (one factor from each form) which have three common items each in the C1S1/MC1S1 match are Factor 3/Factor 2, Factor 3/Factor 3, and Factor 4/Factor 2. Other factor-pairs have fewer than three items in common. These findings suggest that out of the 20 items common to C1S1 and MC1S1[19], 13 items (same items are not recounted) or 65.0% tend to load in the same way.

4.3.2.4 C-TEST 1 SUBTEST 2 (C1S2)

Factor One. Table 4.20 shows that this factor is loaded with five items. These items are #1 (the), #4 (It), #5 (to), #9 (It), and #15 (that). Interestingly, the loadings of these five items are all equal to .995, indicating that the items correlate highly with one another. The items are all high frequency function words. The word it is a repeated item occurring as both #4 and #9. Items #4 and #5 (It has to travel . . .) are adjacent items. The items loading on Factor 1 appear to be related to closed-type items and repeated items.

Factor Two. There are six items loading on this factor. Three of these, #11, #12, and #13, are adjacent items. The other three, #2 (is), #18 (Light), and #24 (travels), are words which also appear unmutilated elsewhere in the test passage. Factor 2 appears to relate to adjacency effect and obvious text clues.

Factor Three The items which load on this factor are #14 (not), #17 (any), and adjacent items #21, #22 and #23. Factor 3 appears to be related to adjacency effect and closed-type items.

Factor Four The items loading on Factor 4 are #2 and #3 (. . . the light is not there instantly.) and #17 (any). These items are all function words. Thus, Factor 4 is

---

[19] Note that even though the number of items in each subtest is equal (=25), the common items for each pair varied, depending upon how many items in the subtest were employed in running factor analysis. While all 25 items in MC1S1 were used, only 20 items in C1S1 were used in the running of factor analysis since the other five items had variances less than .00001, which prevented the computation of factor analysis. (Error 16, Program: FACT20, p. 17).

Table 4.20

Items Loading 0.4 or Greater on Each Factor: Subtest 2 (C1S2 vs. MC1S2)

C1S2				Item	MC1S2				
Factor					Factor				
1	2	3	4		1	2	3	4	5
1				1. the					
	2		4	2. is					5
			4	3. there	1				5
1				4. It				4	
1				5. to				4	
				6. from				4	
				7. flashlight			3		
				8. the		2			
1				9. It		2			
				10. the		2			
	2			11. room		2			5
	2			12. fast	1				
	2			13. you	1	2			
		3		14. not		2	3		
1				15. that		2	3		
				16. is					
		3	4	17. any	1				
	2			18. Light	1	2			
				19. faster	1				
				20. anything	1				
		3		21. that			3		
		3		22. has			3		
		3		23. measured					
	2			24. travels	1				
				25. fast	1				

Note: Items 6, 8, 10, 19, 20, and 25 of C1S2 and items 4 and 16 of MC1S2 were not employed in the analysis.

C-TEST 1

Passage 2 (C1S2)

If you are in a totally dark room and you turn on a flashlight, it will seem that a circle of light appears instantly on the opposite wall. Actually, t_ _ light i_ not th_ _ _ instantly. I_ has t_ travel fr_ _ your flash_ _ _ _ _ to t_ _ wall. I_ crosses t_ _ entire ro_ _ so fa_ _ that y_ _ are n_ _ aware th_ _ it i_ taking a_ _ time.

Li_ _ _ travels fas_ _ _ than anyt_ _ _ _ else th_ _ man h_ _ ever meas_ _ _ _. It tra_ _ _ _ so fa_ _ that the human eye is unable to perceive it move. Scientists have calculated that light travels at a speed of almost three hundred thousand kilometers per second.

MC-TEST 1

Passage 2 (MC1S2)

If you are in a totally dark room and you turn on a flashlight, it will seem that a circle of light appears instantly on the opposite wall. Actually, _ _e light _s not _ _ _re instantly. _t has _o travel _ _om your _ _ _ _ _light to _ _e wall. _t crosses _ _e entire _ _om so _ _st that _ _u are _ _t aware _ _at it _s taking _ _y time.

_ _ _ht travels _ _ _ter than _ _ _ _hing else _ _at man _ _s ever _ _ _ _ured. It _ _ _ _els so _ _st that the human eye is unable to perceive it move. Scientists have calculated that light travels at a speed of almost three hundred thousand kilometers per second.

related to closed-type items.

4.3.2.5 MC-TEST 1 SUBTEST 2 (MC1S2)

Factor One. Table 4.20 reveals that the nine items loading on this factor cluster into three groups of adjacent items (#12 & #13; #17, #18, #19 & #20; and #24 & #25) and one isolated item (#3. there). Factor 1 appears to be mainly related to adjacency effect.

Factor Two. This factor is loaded with eight items. They include two groups of adjacent items (#8, #9, #10, & #11 and #13, #14, & #15) and item #18 (Light), which also appears unmutilated elsewhere in the test passage. The items loading on Factor 2 are involved with adjacency effect and obvious text clues.

Factor Three. The items loading on Factor 3 are comprised of two pairs of adjacent items (#14 & #15 and #21 & #22). The remaining item is #7 (flashlight), which also appears unmutilated in the test passage. This factor appears to be related to adjacency effect and obvious text clues.

Factor Four. This factor is loaded with only one pair of adjacent items #5 & #6 (It has to travel from . . .). Factor 4 is, therefore, related to adjacency effect and closed-type items.

Factor Five. The three items loading on Factor 5 are #2, #3 (. . . the light is not there instantly.), and #11 (room). This factor appears to be involved with adjacency effect and the word room, which also appears unmutilated elsewhere in the test passage.

4.3.2.6 COMPARISON OF FACTOR STRUCTURES (C1S2 & MC1S2)

Table 4.21 shows that in the C1S2/MC1S2 match, the factor pairs which have three or more common items each are Factor 2/Factor 1, Factor 2/Factor 2, and Factor 3/Factor 3. Since same items are not recounted, the total number of common items in

Table 4.21

Cross-tabulation of Items Loading on Each Factor of C1S2 and MC1S2

		MC1S2					
	Factor	1	2	3	4	5	None
C1S2	1		9,15,	15	5		1,4
	2	12,13,18 24	11,13,18			2,11	
	3	17	14	14,21,22			23
	4	3,17				2,3	
	None	19,20,25	8,10	7	6		16

Note:

1) Number of items employed in factor analysis: C1S2 = 19; MC1S2 = 23.
2) Numerals in the cells are test item numbers.

this C1S2/MC1S2 match adds up to eight. As a result, of the 17 common items of C1S2 and MC1S2, eight items or 47.1% appear to load in the same way.

4.3.2.7 C-TEST 1 SUBTEST 3 (C1S3)

Factor One. Table 4.22 shows that there are seven items loading on this factor. These items cluster into a group of five adjacent items (#10, #11, #12, #13, & #14) and two isolated items (#3 & #5). Factor 1 appears to be related mainly to adjacency effect.

Factor Two. There are five items loading on Factor 2. These items are #9 (only), #11 (in), #22 (need), #23 (a), and #25 (only). All the items are words which occur in the test passage more than once. Therefore, this factor is related to obvious text clues and repeated items.

Factor Three. Factor 3 is loaded with five items, namely, #5 (At), #8 (body), #16 (Fortunately), #22 (need), and #23 (a). Three of these items (body, need, and a) are words which also appear unmutilated elsewhere in the test passage. The context for the other two items ("At any one time, . . ." and "Fortunately, it is not difficult . . .") also appears to give adequate cues to the answers. The items which load on Factor 3 are, thus, related to obvious text clues.

Factor Four. The items loading on this factor include #2, #3, #4 (It has enough water to last for several days.), and #13 (last), which is a repeated item. Factor 4 is related to adjacency effect and repeated items.

Factor Five. Factor 5 is loaded with two groups of adjacent items (#6 & #7 and #18, #19, & #20). This factor appears to be an adjacency factor.

4.3.2.8 MC-TEST 1 SUBTEST 3 (MC1S3)

Factor One. There are 12 items loading on Factor 1 (Table 4.22). These items cluster into three groups of adjacent items and a repeated item (#21. oxygen). Factor 1

Table 4.22

Items Loading 0.4 or Greater on Each Factor: Subtest 3 (C1S3 vs. MC1S3)

C1S3					Item	MC1S3			
Factor						Factor			
1	2	3	4	5		1	2	3	4
					1. has				4
			4		2. water		2		
1			4		3. last				4
			4		4. several			3	
1		3			5. At			3	
				5	6. one		2		
				5	7. however		2		
		3			8. body		2		
	2				9. only		2		
1					10. oxygen	1			
1	2				11. in	1			
1					12. lungs	1			
1			4		13. last	1			4
1					14. three	1			
					15. four	1			
		3			16. Fortunately		2		
					17. is	1			
				5	18. difficult	1			
				5	19. us	1	2		
				5	20. acquire		2		
					21. oxygen	1			
	2	3			22. need		2		
	2	3			23. a	1			
					24. we	1			
	2				25. only		2		

Note: Items 1, 17, 21, and 24 of C1S3 were not employed in the analysis.

C-TEST 1

Passage 3 (C1S3)

The body of the average adult has in storage enough food to last for several weeks. It h_ _ enough wa_ _ _ to la_ _ for sev_ _ _ _ days. A_ any o_ _ time, how_ _ _ _, the bo_ _ has on_ _ enough oxy_ _ _ stored i_ the lu_ _ _ to la_ _ for th_ _ _ or fo_ _ minutes! Fortu_ _ _ _ _ _, it i_ not diff_ _ _ _ _ for u_ to acq_ _ _ _ the oxy_ _ _ we ne_ _. As _ rule, w_ need on_ _ to breathe in the air around us for an adequate supply. The amount of oxygen needed by a person depends upon his activities.

MC-TEST 1

Passage 3 (MC1S3)

The body of the average adult has in storage enough food to last for several weeks. It _ _s enough _ _ _er to _ _st for _ _ _ _ral days. _t any _ _e time, _ _ _ _ver, the _ _dy has _ _ly enough _ _ _gen stored _n the _ _ _gs to _ _st for _ _ _ee or _ _ur minutes! _ _ _ _ _ _ately, it _s not _ _ _ _ _cult for _s to _ _ _ _ire the _ _ _gen we _ _ed. As _ rule, _e need _ _ly to breathe in the air around us for an adequate supply. The amount of oxygen needed by a person depends upon his activities.

is, thus, related to adjacency effect.

Factor Two. Factor 2 is loaded with 10 items. These items include two groups of adjacent items (#6, #7, #8, & #9 and #19 & #20) and four isolated items which are mostly content words (an exception is #25. only, which is a repeated item). This factor is, thus, related to adjacency effect and content words.

Factor Three. There are only two items loading on this factor. They are adjacent items #4 and #5 (. . . to last for several days. At any one time, . . .). Factor 3 appears to be related to adjacency effect.

Factor Four. The items loading on this factor are #1 (has); #3 (last), and #13 (last). These items are words which also appear unmutilated elsewhere in the test passage. Factor 4 is, therefore, related to obvious text clues.

4.3.2.9 COMPARISON OF FACTOR STRUCTURES (C1S3 & MC1S3)

Cross-tabulation in Table 4.23 indicates that four factor-pairs in the C1S3/MC1S3 match (Factor 1/Factor 1. Factor 2/Factor 2, Factor 3/Factor 2, and Factor 5/Factor 2) have three or more common items. The results suggest that of the 21 common items of C1S3 and MC1S3, 14 items or 66.7% appear to load similarly.

4.3.2.10 C-TEST 1 SUBTEST 4 (C1S4)

Factor One. Table 4.24 shows that there are nine items loading on Factor 1. These items include a group of five adjaent items (#14, #15, #16, #17, & #18) and isolated items which are mostly content words (#4. group, #9. language, #12. conditions, and #24. other). This factor appears to be related to adjacency effect and content words.

Factor Two. This factor is loaded with eight items. Four of these items cluster into two pairs of adjacent items (#8 & #9 and #21 & #22). The remaining items are mostly function words (#2. this, #12. conditions, #17. with, and #25. however).

Table 4.23

Cross-tabulation of Items Loading on Each Factor of C1S3 and MC1S3

		MC1S3				
	Factor	1	2	3	4	None
C1S3	1	10,11,12 13,14		5	3,13	
	2	11,23	9,22,25			
	3	23	8,16,22	5		
	4	13	2	4	3,13	
	5	18,19	6,7,19 20			
	None	15,17,21 24.			1	

Note:

1) Number of items employed in factor analysis: C1S3 = 21; MC1S3 = 25.
2) Numerals in the cells are test item numbers.

Table 4.24

Items Loading 0.4 or Greater on Each Factor: Subtest 4 (C1S4 vs. MC1S4)

C1S4 Factor 1	C1S4 Factor 2	C1S4 Factor 3	C1S4 Factor 4	C1S4 Factor 5	Item	MC1S4 Factor 1	MC1S4 Factor 2	MC1S4 Factor 3	MC1S4 Factor 4	MC1S4 Factor 5
		3			1. reason				4	
	2			5	2. this			3		
		3			3. that			3		
1					4. group			3		
				5	5. lived		2			
				5	6. isolation		2			
		3			7. reason				4	
	2	3	4		8. that			3	4	
1	2				9. language			3		
			4		10. up	1			4	5*
			4		11. under		2			
1	2				12. conditions		2			
					13. over		2			
1		~~3~~			14. centuries		~~2~~			
1					15. groups	1				
1					16. in	1		3		
1	2		4		17. with	1				
1					18. groups	1		3		
					19. this	1				
			4		20. their	1				
	2				21. grew	1				
	2				22. and					5
			4		23. alike		2			5
1				5	24. other			3		
	2				25. however			3		

* indicates a negative loading.

Note: Item 13 of C1S4 was not employed in the analysis.

C-TEST 1

Passage 4 (C1S4)

The speech of every group of people has developed differently. One rea___ for th__ is th__ each gr___ often li___ in isol_____. Another rea___ is th__ each lang____ grew u_ separately un___ different condi_____. Gradually, ov__ many cent_____, some gro___ came i_ contact wi__ other gro___. When th__ happened, th___ languages gr__ more a__ more al___. In ot___ cases, how____, a group would break apart, and new languages would grow from the original one. The new languages, though distinct in many ways, would maintain similarities to one another.

MC-TEST 1

Passage 4 (MC1S4)

The speech of every group of people has developed differently. One ___son for __is is __at each ___up often ___ed in _____tion. Another ___son is __at each ____uage grew _p separately ___er different _____tions. Gradually, __er many _____ries, some ___ups came _n contact __th other ___ups. When __is happened, ___ir languages __ew more __d more ___ke. In ___er cases, ____ver, a group would break apart, and new languages would grow from the original one. The new languages, though distinct in many ways, would maintain similarities to one another.

Factor 2 appears to be related to adjacency effect and closed-type items.

Factor Three. There are five items loading on this factor. They are #1 (reason), #3 ((that)), #7 (reason), #8 (that), and #13 (centuries). Factor 3 is mainly involved with repeated items.

Factor Four. The items loading on Factor 4 are #8, #10, #11, #17, #20, and #23. All except #23 (alike) are function words. This factor appears to be related mainly to closed-type items.

Factor Five. There are four items loading on this factor. These items are #2 (this), #24 (other), and a pair of adjacent items (. . . each group often lived in isolation.). Factor 5 is involved with adjacency effect and closed-type items.

4.3.2.11 MC-TEST 1 SUBTEST 4 (MC1S4)

Factor One. Table 4.24 shows that seven of the eight items loading on Factor 1 are adjacent items. This factor is apparently related to adjacency effect.

Factor Two. The items loading on this factor include two groups of adjacent items (#5 & #6 and #11, #12, #13, & #14) and #23 (alike). Factor 2 appears to be related to adjacency effect.

Factor Three. There are nine items loading on Factor 3. They cluster into three groups of adjacent items and two isolated items, #16 (in) and #18 (groups). This factor appears to be related mainly to adjacency effect.

Factor Four. Factor 4 is loaded with four items. Three of these items (#1, #7, & #8) are words which occur in the test passage more than once. Factor 4 appears to be related mainly to repeated items.

Factor Five. The items which load on this factor are #10 (up) and a pair of adjacent items (. . . grew more and more alike.). It should be noted that the word up has a negative loading (-0.611), indicating that respondents who answered the adjacent items correctly tended to answer #10 incorrectly, and vice versa. Generally, Facter 5 can be considered as related to adjacency effect.

4.3.2.12 COMPARISON OF FACTOR STRUCTURES (C1S4 & MC1S4)

Table 4.25 reveals that in the C1S4/MC1S4 match, there are four factor-pairs which have at least three common items. The finding indicates that 15 of the 24 common items of C1S4 and MC1S4 or 62.5% appear to demonstrate a similar trend.

4.3.2.13 C-TEST 2 SUBTEST 1 (C2S1)

Factor One. Table 4.26 shows that there are three items loading on Factor 1. These items are #4 (know), #13 (she), and #19 (She). The word she also occurs in the unmutilated first sentence. Item #4 (know) fits here because it is not difficult to figure out. The responses for this item are constrained by the orthographic information (i.e., there are very limited number of words of four letters beginning with kn--). The items loading on Factor 1 appear to be related to obvious text clues.

Factor Two. There are four items loading on Factor 2. These items are #3 (did), #6 (first), #7 (some), and #18 (to). The items are all function words. The adjacent items are #6 and #7 (. . . he first asked some questions, . . .). Factor 2 appears to be related to closed-type items and adjacency effect.

Factor Three. The items loading on this factor are adjacent items #11 and #12 ("How old are you?"), and item #22 (then). Since the word old is highly constrained by the context, Factor 3 can be considered as related to closed-type items and adjacency effect.

Factor Four. The factor is loaded with adjacent items #8, #9, and #10 (. . . and one of them was, . . .), and item #17 (will). The items are all function words. Thus, the items loading on Factor 4 are related to adjacency effect and closed-type items.

Table 4.25

Cross-tabulation of Items Loading on Each Factor of C1S4 and MC1S4

MC1S4

	Factor	1	2	3	4	5	None
C1S4	1	15,16,17 18	12,14	4,9,16 18,24			
	2	17,21	12	2,8,9, 25	8	22	
	3		14	3,8	1,7,8		
	4	10,17,20	12,23	8	8,10	10,23	
	5		5,6	2,24			
	None	19	13				

Note:

1) Number of items employed in factor analysis: C1S4 = 24; MC1S4 = 25.
2) Numerals in the cells are test item numbers.

Table 4.26

Items Loading 0.4 or Greater on Each Factor: Subtest 1 (C2S1 vs. MC2S1)

C2S1				Item	MC2S1					
Factor					Factor					
1	2	3	4		1	2	3	4	5	6
				1. a						
				2. doctor						
	2			3. did				4		
1				4. know					5	
				5. so					5	
	2			6. first			3			6*
	2			7. some			3			
			4	8. and			3			
			4	9. of			3			
			4	10. was			3			
		3		11. old			3			
		3		12. you	1					
1				13. she				4		
				14. I						
				15. remember						
				16. but		2				6
			4	17. will	1					
	2			18. to	1					
1				19. She				4		
				20. for		2				
				21. minute		2				
		3		22. then		2		4		
				23. Yes					5	
				24. remember	1					
				25. doctor	1					6

* indicates a negative loading.

Note: Items 1, 2, 5, 14, 15, 16, 20, 21, 24, and 25 of C2S1 and items 1, 2, 14, and 15 of MC2S1 were not employed in the analysis.

C-TEST 2

Passage 1 (C2S1)

A woman was having some trouble with her heart, so she went to see the doctor. He was _ new doc___, and d__ not kn__ her, s_ he fi___ asked so__ questions, a__ one o_ them w__, "How o__ are y__?"

"Well," s__ answered, "_ don't reme____, doctor, b__ I wi__ try t_ think." S__ thought f__ a min___ and th__ said, "Y__, I reme____ now, doc___! When I married, I was eighteen years old, and my husband was thirty.

MC-TEST 2

Passage 1 (MC2S1)

A woman was having some trouble with her heart, so she went to see the doctor. He was _ new _ _ _tor, and _ _d not _ _ow her, _o he _ _ _st asked _ _me questions, _ _d one _f them _ _s, "How _ _d are _ _u?"

"Well," _ _e answered, "_ don't _ _ _ _mber, doctor, _ _t I _ _ll try _o think." _ _e thought _ _r a _ _ _ute and _ _en said, "_ _s, I _ _ _ _mber now, _ _ _tor! When I married, I was eighteen years old, and my husband was thirty.

4.3.2.14 MC-TEST 2 SUBTEST 1 (MC2S1)

Factor One. Table 4.26 shows that the items loading on Factor 1 are #12 (you), and two pairs of adjacent items #17 & #18 (I will try to think.) and #24 & #25 (. . . I remember now, doctor!). Factor 1 appears to be related to pairs of adjacent words of the same category (i.e., function/function word or content/content word).

Factor Two. The items loading on Factor 2 are #16 (but) and three adjacent items, #20, #21, and #22 (She thought for a minute and then said . . .). Factor 2 seems to be related mainly to adjacency effect.

Factor Three. There are six items loading on Factor 3. These items are #6 through #11. They are adjacent items which are all in the same sentence. The items loading on Factor 3 are obviously related to adjacency effect.

Factor Four. This factor is loaded with four items. These items are #3 (did), #13 (she), #19 (She), and #22 (then), which are all function words. Factor 4 appears to be related to closed-type items.

Factor Five. The items loading on this factor are adjacent items #4 and #5 (. . . not know her, so he . . .) and item #23 (Yes). Factor 5 appears to be mainly related to adjacency effect.

Factor Six. There are three items loading on Factor 6. These items are #6 (first), #16 (but), and #25 (doctor). Item #6 (first) has a negative loading (-0.551), which indicates that the respondent who correctly answered items #16 and #25 would likely incorrectly answer item #6, and vice versa. The items loading on this factor do not seem to be related to one another. Factor 6 appears to be involved with a group of miscellaneous elements which do not seem to share any common characteristic.

4.3.2.15 COMPARISON OF FACTOR STRUCTURES (C2S1 & MC2S1)

There appear to be only 15 common items in C2S1 and MC2S1 (Table 4.27). Factor 4/Factor 3 is the single factor-pair which has three common items. The other factor-pairs have only one or two common items. The findings suggest that, of the 15 common items, only 3 (or 20%) of them seem to load in the same way. This particular subtest of C-Test 2 and MC-Test 2 appears to demonstrate the most dissimilarity with regard to factor structures.

4.3.2.16 C-TEST 2 SUBTEST 2 (C2S2)

Factor One. Table 4.28 shows that there are six items loading on Factor 1. Three items (#14. aware, #17. time, and #18. travels) are content words, while the other three (#12. that, #23. It, and #25. that) are function words which also appear unmutilated in the test passage. This factor is thus related to content words and words which also appear unmutilated.

Factor Two. Factor 2 comprises two groups of adjacent items (#13 & #14 and #20, #21, & #22) and two isolated items (#10. entire and #16. taking). This factor appears to be related to adjacency effect and content words.

Factor Three. There are five items loading on this factor. These items include a pair of adjacent items (#15 & #16), two isolated items (#20. else and #23. It) which are function words, and item #18 (travels) which has a negative loading (-0.562). The items loading on Factor 3 appear to be related to adjacency effect and function words.

Factor Four. Factor 4 is loaded with five items. Three of these items (#3. instantly, #5. travel, and #18. travels) are content words and the other two (#15. It and #22. ever) are function words. All except ever also appear unmutilated elsewhere in the test passage. The items loading on Factor 4, therefore, relate to obvious text clues.

Factor Five. There are three items loading on Factor 5. The factor is related to two content words (#9. crosses and #21. man) and one function word (#4. has).

Table 4.27

Cross-tabulation of Items Loading on Each Factor of C2S1 and MC2S1

	MC2S1							
	Factor	1	2	3	4	5	6	None
C2S1	1				13,19	4		
	2	18		6,7	3		6	
	3	12	22	11	22			
	4	17		8,9,10				
	None	24,25	16,20,21			5,23	16,25	1,2,14 15

Note:

1) Number of items employed in factor analysis: C2S1 = 15; MC2S1 = 21.
2) Numerals in the cells are test item numbers.

Table 4.28

Items Loading 0.4 or Greater on Each Factor: Subtest 2 (C2S2 vs. MC2S2)

C2S2						Item	MC2S2				
Factor							Factor				
1	2	3	4	5	6		1.	2	3	4	5
						1. light					
					6	2. not	1				
			4			3. instantly				4	
				5		4. has	1				
			4			5. travel			3		
						6. your					5
						7. to					5
						8. wall					5
				5		9. crosses	1				
	2					10. entire				4	
						11. so				4	
1						12. that		2			
	2					13. are	1				
1	2					14. aware	1				
		3	4			15. it		2	3		
	2	3				16. taking	1				
1						17. time	1				
1		3•	4			18. travels			3		
					6	19. than		2			
	2	3				20. else			3		
	2			5		21. man	1				
	2		4			22. ever	1		3		
1		3				23. It			3		
						24. so				4	
1						25. that		2			

• indicates a negative loading.

Note: Items 1, 6, 7, 8, 11, and 24 of C2S2 and item 1 MC2S2 were not employed in the analysis.

C-TEST 2

Passage 2 (C2S2)

If you are in a totally dark room and you turn on a flashlight, it will seem that a circle of light appears instantly on the opposite wall. Actually, the li_ _ _ is n_ _ there inst_ _ _ _ _. It h_ _ to tra_ _ _ from yo_ _ flashlight t_ the wa_ _. It cro_ _ _ _ the ent_ _ _ room s_ fast th_ _ you a_ _ not aw_ _ _ that i_ is tak_ _ _ any ti_ _.

Light tra_ _ _ _ faster th_ _ anything el_ _ that m_ _ has ev_ _ measured. I_ travels s_ fast th_ _ the human eye is unable to perceive it move. Scientists have calculated that light travels at a speed of almost three hundred thousand kilometers per second.

MC-TEST 2

Passage 2 (MC2S2)

If you are in a totally dark room and you turn on a flashlight, it will seem that a circle of light appears instantly on the opposite wall. Actually, the _ _ _ht is _ _t there _ _ _ _ _ntly. It _ _s to _ _ _vel from _ _ur flashlight _o the _ _ll. It _ _ _ _ses the _ _ _ire room _o fast _ _at you _ _e not _ _ _re that _t is _ _ _ing any _ _me.

Light _ _ _ _els faster _ _an anything _ _se that _ _n has _ _er measured. _t travels _o fast _ _at the human eye is unable to perceive it move. Scientists have calculated that light travels at a speed of almost three hundred thousand kilometers per second.

Factor 5 seems to involve open-type items. The word has belongs to this factor because it permits other responses (e.g., had, hit).

Factor Six. Factor 6 is loaded with only two items (#2. not and #19. than). This factor appears to be related to closed-type items.

4.3.2.17 MC-TEST 2 SUBTEST 2 (MC2S2)

Factor One. There are nine items loading on Factor 1 (Table 4.28). These items cluster into three pairs of adjacent items and three isolated ones (#2. not, #4. has, and #9. crosses). Factor 1 appears to be mainly related to adjacency effect.

Factor Two. Factor 2 is loaded with items #12 (that), #15 (it), #19 (than), and #25 (that), which are all function words. This factor, thus, relates to closed-type items.

Factor Three. There are six items loading on this factor. The items include four function words and items #5 (travel) and #18. (travels). The items loading on Factor 3 appear to be related to function words and the word travel(s), which also appears unmutilated in the test. Factor 3 is, therefore, involved with closed-type items and obvious text clues.

Factor Four. This factor is loaded with four items of which two (#10 & #11) are adjacent. These four items comprise two content words (#3. instantly and #10. entire) and two occurrences of the word so (#11 & #24). Factor 4 appears to be related to content words and repeated items.

Factor Five. The items loading on Factor 5 are #6, #7, and #8 (It has to travel from your flashlight to the wall.). This factor is obviously related to adjacency effect.

4.3.2.18 COMPARISON OF FACTOR STRUCTURES (C2S2 & MC2S2)

Cross-tabulation in Table 4.29 reveals that four factor-pairs of the C2S2/MC2S2 match have three or more common items. Generally, of the 19 common items of C2S2 and MC2S2, 10 items (or 52.6%) appear to load in a like manner.

4.3.2.19 C-TEST 2 SUBTEST 3 (C2S3)

Factor One. Table 4.30 shows that there are five items loading on Factor 1. Examination of the table reveals that the items are comprised of two occurrences each of the words enough (#1 & #9) and the (#7 & #20), and item #6 (time) which is adjacent to #7 (the). Factor 1 appears to be mainly related to repeated items which in fact also appear unmutilated in the test passage.

Factor Two. Factor 2 is loaded with a pair of adjacent items (As a rule . . .) and items #5 (any) and #15 (minutes). The latter two items belong to this factor probably because of the contextualconstraint ("At any one time . . ." and ". . . to last for three or four minutes!"). This factor is, thus, involved mainly with adjacency effect and closed-type items.

Factor Three. There are three items loading on this factor. These items include two occurrences of the (#7 & #11) and #14(or). Factor 3 appears to be related to closed-type items.

Factor Four. This factor is loaded with two content words (#6. time and #10. stored) and two function words (#11. the and #17. not). Note that the former two items both have negative loadings. Factor 4 is, therefore, invoved with a negative relationship between content words and function words.

Factor Five. There are only two words (#8. has and #25. to) loading on Factor 5. This factor appears to be related to closed-type items.

Table 4.29

Cross-tabulation of Items Loading on Each Factor of C2S2 and MC2S2

MC1S2

	Factor	1	2	3	4	5	None
C2S2	1	14,17	12,25	18,23			
	2	13,14,16 21,22		20,22	10		
	3	16	15	15,18,20 23			
	4	22	15	5,15,18 22	3		
	5	4,9,21					
	6	2	19				
	None				11,24	6,7,8	1

Note:

1) Number of items employed in factor analysis: C2S2 = 19; MC2S2 = 24.
2) Numerals in the cells are test item numbers.

Table 4.30

Items Loading 0.4 of Greater on Each Factor: Subtest 3 (C2S3 vs. MC2S3)

C2S3					Item	MC2S3					
Factor						Factor					
1	2	3	4	5		1	2	3	4	5	6
1					1. enough						
					2. to	1					
					3. for						6
					4. days		2				
	2				5. any		2				
1			4*		6. time		2				
1		3			7. the				4		
				5	8. has			3			
1					9. enough	1					
			4*		10. stored		2	3			
		3	4		11. the				4	5	
					12. to	1				5	
					13. for	1					6
		3			14. or	1					
	2				15. minutes			3			
					16. it						
			4		17. not		2		4		
					18. for	1					
					19. to						
1					20. the				4		
					21. we			3			
	2				22. As		2				
	2				23. rule		2				
					24. need					5	
				5	25. to					5	

* indicates a negative loading.

Note: Items 2, 3, 4, 12, 13, 16, 18, 19, 21, and 24 of C2S3 and items 1, 16, and 19 of MC2S3 were not employed in the analysis.

C-TEST 2

Passage 3 (C2S3)

The body of the average adult has in storage enough food to last for several weeks. It has eno___ water t_ last f__ several da__. At a__ one ti__, however, t__ body h__ only eno___ oxygen sto___ in t__ lungs t_ last f__ three o_ four min____! Fortunately, i_ is n__ difficult f__ us t_ acquire t__ oxygen w_ need. A_ a ru__, we ne__ only t_ breathe in the air around us for an adequate supply. The amount of oxygen needed by a person depends upon his activities.

MC-TEST 2

Passage 3 (MC2S3)

The body of the average adult has in storage enough food to last for several weeks. It has _ _ _ugh water _o last _ _r several _ _ys. At _ _y one _ _me, however, _ _e body _ _s only _ _ _ugh oxygen _ _ _red in _ _e lungs _o last _ _r three _r four _ _ _ _tes! Fortunately, _t is _ _t difficult _ _r us _o acquire _ _e oxygen _e need. _s a _ _le, we _ _ed only _o breathe in the air around us for an adequate supply. The amount of oxygen needed by a person depends upon his activities.

4.3.2.20 MC-TEST 2 SUBTEST 3 (MC2S3)

Factor One. There are six items loading on Factor 1 (Table 4.30). These items cluster into a group of three adjacent items (#12, #13, & #14) and three isolated ones (#2, #9, & #18). The word enough (#9), which is the only one that is not a function word, also appears unmutilated in the test passage. Factor 1 is, thus, related to adjacency effect and closed-type items.

Factor Two. Factor 2 is loaded with two groups of adjacent items (#4, #5, & #6 and #22 & #23) and items #10 (stored) and #17 (not). This factor appears to relate mainly to adjacency effect.

Factor Three. This factor is loaded with items #8 (has), #10 (stored, #15. (minutes), and #21. (we). The word we though the only item that is not a content word, allows for other responses, (e.g., he, me). Factor 3 is, therefore, related to open-type items.

Factor Four. Four items load on this factor. Interestingly, three of these items (#7, #11, & #20) are three occurrences of the word the. The remaining item is #17 not. Factor 4 appears to be related to closed-type items and repeated items.

Factor Five. Factor 5 is loaded with two pairs of adjacent items (#11, & #12 and #24 & #25). This factor is, thus, related to adjacency effect.

Factor Six. The items loading on this factor are #3 (for) and #13 (for). Factor 6 appears to be involved with closed-type and repeated items.

4.3.2.21 COMPARISON OF FACTOR STRUCTURES (C2S3 & MC2S3)

Table 4.31 shows that two factor-pairs (Factor 2/Factor 2 & Factor 4/Factor 2) of the C2S3/MC2S3 match have three common items each. Superficially, it seems that 6 of the 14 common items in C2S3 and MC2S3 demonstrate a similar loading structure. A close examination, however, reveals that the loadings of items #6 and #10 of C2S3 both

Table 4.31

Cross-tabulation of Items Loading on Each Factor of C2S3 and MC2S3

		MC2S3						
	Factor	1	2	3	4	5	6	None
	1	9	6		7,20			1
	2		5,22,23	15				
C2S3	3	14			7,11	11		
	4		6,10,17	10	11,17	11		
	5			8		25		
	None	2,12,13 18	4	21		12,24	3,13	16,19

Note:

1) Number of items employed in factor analysis: C2S3 = 15; MC2S3 = 22.
2) Numerals in the cells are test item numbers.

have negative values, while their loadings in MC2S2 show a positive ones; thus, they do not load in the same direction in the two formats. As a consequence, it is likely that only 4 of the 14 common items (or 28.6%) load in the same way.

4.3.2.22 C-TEST 2 SUBTEST 4 (C2S4)

Factor One. Table 4.32 shows that there are seven items loading on Factor 1. These items include two groups of adjacent items (#17 & #18 and #20, #21, & #22), and #13 (many) and #25 (a). The items loading on Factor 1 are related to adjacency effect and closed-type items.

Factor Two. This factor is loaded with nine items which cluster into three groups of adjacent items and items #6 (Another) and #19 (happened). Factor 2 appears to be related mainly to adjacency effect.

Factor Three. There are six items loading on this factor. The items include a group of adjacent items (. . . each group often lived in isolation) and items #8 (each), #11 (different), and #20 (languages). The latter three are words which occur in the test passage more than once (with #11 occurring unmutilated in an adverbial form: differently). The items loading on Factor 3 are related to adjacency effect and obvious text clues.

Factor Four. Factor 4 is loaded with seven items. These items consist of a pair of adjacent items (#11 & #12) and five function words. Factor 4 appears to be related to closed-type items.

4.3.2.23 MC-TEST 2 SUBTEST 4 (MC2S4)

Factor One. Table 4.32 shows that there are 10 items loading on Factor 1. Eight of these items (#16 through #23) are adjacent items. Factor 1 is related mainly to adjacency effect.

Table 4.32

Items Loading 0.4 or Greater on Each Factor: Subtest 4 (C2S4 vs. MC2S4)

C2S4				Item	MC2S4					
Factor					Factor					
1	2	3	4		1	2	3	4	5	6
				1. for						6
			4	2. is						
		3		3. each				4		
		3		4. often					5	
		3	4	5. in			3			
	2			6. Another		2				
			4	7. Is			3			
		3		8. each				4		
	2			9. grew			3			
	2			10. separately		2				
		3	4	11. different		2				
			4	12. Gradually		2				
1				13. many	1		3			
	2			14. some			3			
	2			15. came		2				
	2			16. contact	1	2				
1				17. other	1					
1			4	18. When	1	2				
	2			19. happened	1					
1		3		20. languages	1					
1				21. more	1				5	
1				22. more	1				5	
	2			23. In	1					6
	2			24. cases						
1			4	25. a	1					

Note: Item 1 of C2S4 and item 2 of MC2S4 were not employed in the analysis.

C-TEST 2

Passage 4 (C2S4)

The speech of every group of people has developed __ differently. One reason f_ _ this i__ that ea_ _ group of_ _ _ lived i_ isolation. Ano_ _ _ _ reason i_ that ea_ _ language gr_ _ up separ_ _ _ _ _ under diff_ _ _ _ _ conditions. Grad_ _ _ _ _, over ma_ _ centuries, so_ _ groups ca_ _ in con_ _ _ _ with ot_ _ _ groups. Wh_ _ this happ_ _ _ _, their lang_ _ _ _ _ grew mo_ _ and mo_ _ alike. I_ other ca_ _ _, however, _ group would break apart, and new languages would grow from the original one. The new languages, though distinct in many ways, would maintain similarities to one another.

MC-TEST 2

Passage 4 (MC2S4)

The speech of every group of people has developed differently. One reason _ _r this _s that _ _ch group _ _ _en lived _n isolation. _ _ _ _her reason _s that _ _ch language _ _ew up _ _ _ _ _ately under _ _ _ _ _rent conditions. _ _ _ _ _ally, over _ _ny centuries, _ _me groups _ _me in _ _ _ _act with _ _ _er groups. _ _en this _ _ _ _ened, their _ _ _ _ _ages grew _ _re and _ _re alike. _n other _ _ _es, however, _ group would break apart, and new languages would grow from the original one. The new languages, though distinct in many ways, would maintain similarities to one another.

Factor Two. This factor is loaded with two groups of adjacent items (#10, #11, & #12 and #15 & #16) and two function words (#6. Another and #18. When). The items loading on Factor 2 appear to be related to adjacency effect and two closed-type items.

Factor Three. Five items load on Factor 3. These items include #5 (in), #7 (is), #9 (grew), and a pair of adjacent items (#13 & #14). Factor 3 appears to be mainly related to closed-type items.

Factor Four. Factor 4 is loaded with two occurrences of the word each (items #3 & #8). This factor is, thus, related to repeated items.

Factor Five. The items loading on Factor 5 include #4 (often), #21 (more), and #22 (more). This factor is, therefore, related to closed-type and repeated items.

Factor Six. This factor is loaded with items #1 (for) and #23 (In), which are both function words. Factor 6 appears to be related to closed-type items.

4.3.2.24 COMPARISON OF FACTOR STRUCTURES (C2S4 & MC2S4)

Cross-tabulation of items which load on each factor of C2S4 and MC2S4 in Table 4.33 shows that four factor-pairs have three or more common items. An investigation reveals that 15 items are included in these factor-pairs. As a result, it appears that of the 23 common items of C2S4 and MC2S4, 15 items or 65.2% appear to load in the same way.

4.3.3 SUMMARY OF FACTOR ANALYSIS

The 80 factors extracted by the 16 analyses can be identified as having five common characteristics:

A. Adjacency effect;

B. Closedness of items (characteristic mostly of function words);

Table 4.33

Cross-tabulation of Items Loading on Each Factor of C2S4 and MC2S4

MC2S4

	Factor	1	2	3	4	5	6	None
C2S4	1	13,17,18 20,21,22 25	18	13		21,22		
	2	16,19,23	6,10,15 16	9,14			23	24
	3	20	11	5	3,8	4		
	4	18,25	11,12,18	5,7				2
	None						1	

Note:

1) Number of items employed in factor analysis: C2S4 = 24; MC2S4 = 24.
2) Numerals in the cells are test item numbers.

C. Openness of items (characteristic mostly of content words);

D. Obvious text clues (words which also appear unmutilated elsewhere in the test passage); and

E. Practice effect (repeated items).

Cross-comparison of test items in the two test formats can be summarized as follows:

1. Generally, the C-Test and the correspondent MC-Test do not have the same underlying structure. Of the 153 items compared, only 82 items (53.6%) appear to load in a similar way.
2. Adjacency effect is a prominent characteristic in most subtests. Of the 80 factors extracted, 46 (57.5%) are related to adjacency effect.
3. Adjacency effect appears to be more salient in the MC-Test than in the C-Test. Whereas 61.9% (26 out of 42) of the factors in the MC-Test are involved with adjacency effect, 52.6% (20 out of 38) of the factors in the C-Test are related to adjacency effect.
4. Of the 80 factors, 32 (40.0%) have a characteristic of item closedness and 10 (12.5%) relate to item openness.
5. Thirteen factors (eight from the C-Test and five from the MC-Test) are found to be related to "obvious text clues", words which also appear unmutilated elsewhere in the test passage.
6. There are 12 factors (five from the C-Test and seven from the MC-Test) which are involved with repeated items.
7. When mutilations begin at the third words (C-Test 2 & MC-Test 2) instead of the second word of the second sentence (C-Test 1 & MC-Test 1), the same reading passages appear to produce tests which have different factor structures. In the C-Test 1/MC-Test 1 pair, 61.0% of the items appear to load in the same way, whereas in the C-Test 2/MC-Test 2 pair, only 45.1% of the items demonstrate the same trend.

Thus, it is evident that the C-Test has a different factor structure from the corresponding MC-Test which has been constructed from the same reading passages. The MC-Test demonstrates to be a little more heavily involved with adjacency factors than the

C-Test, though adjacency effect is predominant in both test formats. This suggests that when the first part of the word is deleted (the MC-Test), the degree of interdependence of items tends to be higher than when the second part is deleted (the C-Test). In other words, when the second part of the word is present, the correct (or incorrect) answer to one item tends to have more impact on the answer to the adjacent item than when the first part is present, implying that in the MC-Test, the completion of an item is based more on the completed-item context rather than solely on the orthographic information available in the item.

More factors appear to relate to item closedness than item openness. This is because, as mentioned in Part I (4.1.2) of this chapter, there are more function-word items than content-word items in all test forms, and some of the latter also appear to be relatively "closed", due to the contextual constraint.

Since the number of factors relating to words which also appear unmutilated elsewhere in the test passage is greater in the C-Test than in the MC-Test, it suggests that these words are generally easier to recognize when the first part is present than when the second part is present. However, for the 12 factors interpreted as involving repeated items, more of this characteristic is found in the MC-Test than in the C-Test, which implies that in the former the respondents would need to check back the answer previously completed due to the difficulty of items.

Different starting points of mutilation appear to have an impact on the underlying structures of the tests. In this study, when the mutilation begins at the second word of the second sentence (C-Test 1/MC-Test 1), the percentage of items that load in the same way is higher than when the mutilation begins at the third word (C-Test 2/MC-Test 2).

As far as literature is concerned, there has not been any research conducted using factor analysis to establish the construct validity of the C-Test in this manner. The uses of factor analysis as reported in Raatz & Klein-Braley (1981), Raatz (1984), and Klein-Braley & Raatz (1984), though all purporting to establish the "factorial" validity of the C-Tests, were largely to determine the intercorrelation between the C-Tests and the other tests and/or criteria, rather than to establish their factor structures.[20]

The only construct validity study using factor analysis known that is comparable to this investigation is the research conducted by Lee (1985). Lee attempted to establish the validity of the cloze test items. The cloze tests investigated were comprised of three passages of different subject matter, employing every seventh word deletion, 35 blanks each. The subjects were 146 freshmen at the University of Hong Kong. The results of the component analysis showed that only one factor emerged as significant. This factor was bipolar,[21]and further analyses indicated that these cloze tests measured "some sort of overall language ability" which seemed to correspond to an "openness" versus "closedness" opposition of the test items.

Due to the different item structures of the C-Test/MC-Test and the every seventh word deletion cloze test, the factors abstracted in this study are not the same as those found in Lee's. The factors are not bipolar. In addition, adjacency effect is found to be predominant, although the openness/closedness opposition is quite distinct. The other two factors, obvious text clues and repeated items, however, could be considered as special characteristics of item closedness.

4.3.4 CORRELATIONAL ANALYSIS (L2 DATA)

Another aspect of validity investigated is criterion-related validity. The information available for this correlational investigation is from L2 data. Table 4.34 displays Pearson product moment coefficients between the Michigan Test scores and the C-Test/MC-Test scores.

As shown in Table 4.34, the correlation coefficients range from .423 to .606. Comparing to the results summarized in Klein-Braley & Raatz (1984), which reported that the correlations between the C-Test and the DELTA (Duisburg English Language Test for

---

[20] The results of these studies all suggested that the C-Tests loaded highly together with other tests and/or criteria which were known as measures of global language ability, implying that the C-Tests possibly measure general language ability.

[21] A bipolar factor is one that has both positively and negatively loaded variables.

Table 4.34

Correlations between the Michigan Test and the C-Test/MC-Test

	Michigan Test	Probability	n
C-Test 1	0.542	0.008	23
MC-Test 1	0.423	0.056	21
C-Test 2	0.606	0.006	19
MC-Test 2	0.442	0.051	20

Advanced Students) of the nine investigations ranged from .62 to .90, the correlations in the present study (especially, the Michigan Test and the MC-Test) are fairly low.

It was expected that the correlation between the MC-Test and the Michigan Test would be higher than the correlation between the C-Test and the Michigan Test, because linguistically the MC-Test seems to measure language ability better than the C-Test. Furthermore, since MC-Test items discriminated better than C-Test items, it would seem that the MC-Test/Michigan Test correlation would be higher. Unfortunately, the data did not bear this out. Perhaps this is mere sampling variation since the subjects involved in the C-Test/Michigan Test correlation was not the same as the sample on which the MC-Test/Michigan Test correlation was calculated.

4.3.5 SUMMARY OF CORRELATIONAL ANALYSIS

The findings from the correlational analyses between the Michigan Test and the C-Test/MC-Test suggest that there is a certain amount of relationship and that they share significant characteristics inspite of the different physical test structures. Of the two corresponding formats, the C-Test appears to predict the performance on the Michigan Test better than the MC-Test for this L2 sample. Since these correlations were based on a small number of data points(i.e., between 19 and 23 subjects), the results could have been affected by sampling variation.

4.4 ERROR ANALYSIS AND INTERVIEWS

In order to outline the error patterns of the test responses for L2 data, an error analysis was conducted on the test responses of 28 (or 26.9%) of the 104 ESL students. An equal number of students were randomly chosen from the High Proficiency Group (HPG or the first quarter) and the Low Proficiency Group (LPG or the fourth quarter), as determined by the Michigan Test composite scores.[22] There were 14 students in each group. Of the 28 students, 12 took the C-Test format and 16 took the MC-Test format. Within one week after the administration of the written tests, all of the 28 students were individually interviewed by the researcher to determine their reading strategies. These interviews were conducted in a quiet room where tape-recording equipment had been arranged. All the interviews were then tape-recorded for later determination of the readers' strategies and their opinions of this type of test.

---

[22] The ranges of the Michigan Test composite scores are from 52 to 96 for the HPG and from 15 to 30 for the LPG.

4.4.1 ERROR ANALYSIS

4.4.1.1 C-TEST ERROR INVENTORY

C-Test Error Inventory (CEI). The CEI coding sheet was devised by the researcher. The errors were initially classified into three categories: Substitution, No Response, and Unintelligible. The Substitution errors were further investigated through two questions adapted from Goodman & Burke's (1972) Reading Miscue Inventory questions. These categories have been chosen because they seemed to be most relevant to the nature of the data.

The following are the three categories of errors.

1. Substitution. An error is treated as a "Substitution" error only when the respondent supplied a word which was not the original word but was nevertheless interpretable. Both of the following questions are asked about that word.
    1.1 Grammatical Function. Is the grammatical function of the error the same as the grammatical function of the expected response?
    1.2 Meaning Change. To what extent does the error affect the meaning of the text?
2. No Response. When the respondent did not supply a word for the blank, the error is classified as a "No Response" error and questions 1.1 and 1.2 are automatically discarded.
3. Unintelligible. When the respondent supplied a word that appears to be

unidentifiable or uninterpretable, the error is classified as an "Unintelligible" error (cf., Substitution), and questions 1.1 and 1.2 are automatically discarded.

All the errors of each of the 28 interviewees were written on separate C-Test coding sheets and coded according to the criteria mentioned above (see examples in Appendix E). The analysis of errors was done by the researcher. For errors which seemed ambiguous, two English language specialists were independently consulted. When disagreement occurred, the researcher made the final decision.

4.4.1.2 OVERALL RESULTS

The overall results of the error analysis are reported in Tables 4.35 and 4.36. The findings in Table 4.35 reveal that on the 100 items, the 28 students made a total of 820 errors, or an average of 29.3 errors per student. The LPG made approximately twice as many errors as the HPG, (545:275). The percentage of errors made by the HPG is highest with the Substitution errors (54.9%), whereas the highest percentage of errors in the LPG is with the No Response errors (48.4%). The percentage of the Unintelligible errors is higher in the LPG than in the HPG (6.4%:2.2%). The findings thus indicate that the errors made by the high proficiency students tend to be those more easily interpreted in terms of grammatical function and meaning than those made by the low proficiency students.

The results given in Table 4.36 demonstrate that the highest percentages of the errors found are Substitution (66.3%) for the C-Test and No Response (57.4%) for the MC-Test. The percentages of the Unintelligible errors are 6.5% for the C-Test and 4.2% for the MC-Test. The data suggest that when the first part of the word is given (the C-Test), the respondents are more tempted to answer than when the last part of the word is given (the MC-Test). Therefore, when the respondents come to a difficult item, they tend to try to find a word that may fit the blank in the C-Test format but tend to give up (i.e., leave blanks) in the MC-Test format because of insufficient information.

Table 4.35

Errors Identified in High and Low Proficiency Groups

Proficiency Group	Number of Students	Error Total	Error Type		
			Substitution	No Response	Unintelligible
High	14	275	151 (54.9%)	118 (42.9%)	6 (2.2%)
Low	14	545	246 (45.1%)	264 (48.4%)	35 (6.4%)
Total	28	820	397 (48.4%)	382 (46.6%)	41 (5.0%)

Table 4.36

Errors Identified in C-Test and MC-Test Formats

Format	Number of Students	Error Total	Error Type		
			Substitution	No Response	Unintelligible
C-Test	12	294	195 (66.3%)	80 (27.2%)	19 (6.5%)
MC-Test	16	526	202 (38.4%)	302 (57.4%)	22 (4.2%)
Total	28	820	397 (48.4%)	382 (46.6%)	41 (5.0%)

4.4.1.3 SUBSTITUTION ERRORS

The Substitution errors have been further examined in reference to their grammatical functions and meanings. The results are displayed in Talbes 4.37 and 4.38.

Table 4.37 illustrates that with regard to grammatical function, only 31.2% of the 397 Substitution errors have the same grammatical function as the expected responses. With respect to meaning, as low as 8.8% of all the Substitution errors retain the overall intended meaning. The errors made by the HPG show more favorable percentages in relation to both the similarity of grammatical function and the retention of meaning than the errors made by the LPG. The results thus suggest that the high proficiency students may have made more use of both syntactic/structural and semantic information available in the text than the low proficiency students did.

As can be seen in Table 4.38, there does not seem to be much difference between the percentages of the errors in the two test formats with regard to the similarity of grammatical function. Predictably, the percentage is higher in the MC-Test format than in the C-Test format, because there is normally more syntactic/structural information available in the former than in the latter. In the retention of overall intended meaning, however, more of the errors in the C-Test format tend to retain the original meaning than those in the MC-Test format. Again, this is not surprising because, as has been discussed earlier, there tends to be more semantic information in the C-Test items than in the MC-Test items.

In order to make a detailed comparison, the number of errors along with the common error for each of the items in all the subtests have been displayed in Tables 4.39 through 4.46.

The data in Tables 4.39 to 4.46 reveal that the majority of common errors[23] found in the MC-Test format belong to the "No Response" category. This is not so

---

[23] Note that because the error analysis has been performed on only the test responses of the 28 students chosen for interviewing, the number of students taking each form of the test is only 6 for each C-Test form and 8 for each MC-Test form. Thus, an error is considered a "common error" when two or more students made the same mistake.

Table 4.37

Substitution Errors Identified in High and Low Proficiency Groups

Proficiency Group	Number of Students	Error Total	Grammatical Function		Meaning Change	
			Same	Different	Little or No	High
High	14	151	66 (43.7%)	85 (56.3%)	16 (10.6%)	135 (89.4%)
Low	14	246	58 (23.6%)	188 (76.4%)	19 (7.7%)	227 (92.3%)
Total	28	397	124 (31.2%)	273 (68.8%)	35 (8.8%)	362 (91.2%)

Table 4.38

Substitution Errors Identified in C-Test and MC-Test Formats

Proficiency Format	Number of Students	Error Total	Grammatical Function		Meaning Change	
			Same	Different	Little or No	High
C-Test	12	195	57 (29.2%)	138 (70.8%)	33 (16.9%)	162 (83.1%)
MC-Test	16	202	67 (33.2%)	135 (66.8%)	2 (1.0%)	200 (99.0%)
Total	28	397	124 (31.2%)	273 (68.8%)	35 (8.8%)	362 (91.2%)

Table 4.39

Comparison of Errors for C1S1 and MC1S1

Item	Errors		Common Error	
	C1S1	MC1S1	C1S1	MC1S1
1. was	0	1	-	-
2. new	3	6	-	Now, (no response)
3. and	1	2	-	-
4. not	1	2	-	-
5. her	3	2	how	-
6. he	0	1	-	-
7. asked	1	5	-	(no response)
8. questions	0	5	-	(no response)
9. one	1	4	-	she, (no response)
10. them	3	1	that	-
11. How	0	0	-	-
12. are	0	0	-	-
13. Well	1	3	-	(no response)
14. answered	1	4	-	(no response)
15. don't	0	1	-	-
16. doctor	0	1	-	-
17. I	0	0	-	-
18. try	2	4	-	say
19. think	2	3	-	(no response)
20. thought	0	2	-	-
21. a	0	0	-	-
22. and	2	0	are	-
23. said	2	0	-	-
24. I	0	0	-	-
25. now	1	4	-	how

Note: C1S1, N=6; MC1S1, N=8.

Table 4.40

Comparison of Errors for C1S2 and MC1S2

Item	Errors		Common Error	
	C1S2	MC1S2	C1S2	MC1S2
1. the	1	3	-	see
2. is	1	0	-	-
3. there	2	6	-	aware, (no response)
4. It	0	0	-	-
5. to	1	2	-	-
6. from	2	3	-	room
7. flashlight	0	0	-	-
8. the	1	3	-	see
9. It	2	1	Is	-
10. the	1	3	-	see
11. room	2	0	-	-
12. fast	1	2	-	-
13. you	0	1	-	-
14. not	2	2	now	-
15. that	1	2	-	-
16. is	0	0	-	-
17. any	4	1	are	-
18. light	1	0	-	-
19. faster	0	2	-	(no response)
20. anything	1	3	-	(no response)
21. that	2	3	-	what
22. has	4	3	how	was
23. measured	3	6	-	(no response)
24. travels	0	1	-	-
25. fast	1	2	-	-

Note: C1S2, N=6; MC1S2, N=8.

Table 4.41

Comparison of Errors for C1S3 and MC1S3

Item	Errors		Common Error	
	C1S3	MC1S3	C1S3	MC1S3
1. has	2	3	-	was
2. water	1	5	-	(no response)
3. last	2	1	-	-
4. several	1	1	-	-
5. At	3	2	-	it
6. one	3	5	off	the, (no response)
7. however	1	3	-	(no response)
8. body	1	1	-	-
9. only	3	5	once	(no response)
10. them	0	2	-	-
11. in	4	1	is	-
12. lungs	5	5	(no response)	(no response)
13. last	4	3	-	(no response)
14. three	1	3	-	(no response)
15. four	0	3	-	(no response)
16. Fortunately	0	7	-	(no response)
17. is	0	1	-	-
18. difficult	3	5	different	(no response)
19. us	1	1	-	-
20. acquire	4	8	-	(no response)
21. oxygen	0	2	-	-
22. need	0	2	-	-
23. a	0	2	-	(no response)
24. we	0	3	-	-
25. only	3	5	once	(no response)

Note: C1S3, N=6; MC1S3, N=8.

Table 4.42

Comparison of Errors for C1S4 and MC1S4

Item	Errors		Common Error	
	C1S4	MC1S4	C1S4	MC1S4
1. reason	1	5	-	person
2. this	3	2	that	(no response)
3. that	3	1	then	-
4. group	1	1	-	-
5. lived	5	7	lives	(no response)
6. isolation	3	8	(no response)	(no response)
7. reason	1	5	-	person
8. that	2	2	-	(no response)
9. language	0	0	-	-
10. up	3	1	us	-
11. under	2	8	(no response)	other, (no response)
12. conditions	2	8	-	(no response)
13. over	0	6	-	(no response)
14. centuries	4	8	-	countries, (no response)
15. groups	1	2	-	(no response)
16. in	3	0	is	-
17. with	1	1	-	-
18. groups	2	1	-	-
19. this	4	3	that	itis
20. their	5	2	-	(no response)
21. grew	4	3	grow	-
22. and	3	1	are	-
23. alike	5	8	alway, (no response)	(no response)
24. other	2	4	-	(no response)
25. however	1	3	-	(no response)

Note: C1S4, N=6; MC1S4, N=8.

Table 4.43

Comparison of Errors for C2S1 and MC2S1

Item	Errors		Common Error	
	C2S1	MC2S1	C2S1	MC2S1
1. a	0	0	-	-
2. doctor	0	0	-	-
3. did	0	1	-	-
4. know	1	4	-	show
5. so	0	1	-	-
6. first	3	3	(no response)	(no response)
7. some	1	3	-	tome
8. and	1	3	-	did
9. of	3	1	-	-
10. was	4	2	(no response)	-
11. old	0	3	-	did
12. you	0	0	-	-
13. she	1	2	-	are
14. I	1	0	-	-
15. remember	0	0	-	-
16. but	1	3	-	-
17. will	2	0	-	-
18. to	0	0	-	-
19. She	0	0	-	-
20. for	3	2	few	-
21. minute	1	2	-	(no response)
22. then	2	2	they	-
23. Yes	3	3	You	-
24. remember	0	0	-	-
25. doctor	0	0	-	-

Note: C2S1, N=6; MC2S1, N=8.

Table 4.44

Comparison of Errors for C2S2 and MC2S2

Item	Errors		Common Error	
	C2S2	MC2S2	C2S2	MC2S2
1. light	0	0	-	-
2. not	0	4	-	-
3.instantly	0	2	-	-
4. has	3	5	-	was
5. travel	1	2	-	-
6. your	2	5	-	(no response)
7. to	0	0	-	-
8. wall	0	0	-	-
9. crosses	2	5	-	(no response)
10. entire	4	5	-	(no response)
11. so	0	6	-	to
12. that	2	4	then	what, (no response)
13. are	0	4	-	see
14. aware	3	7	-	there, (no response)
15. it	0	1	-	-
16. taking	1	7	-	(no response)
17. time	1	5	-	(no response)
18. travels	0	2	-	-
19. than	4	1	then, that	-
20. else	0	4	-	(no response)
21. man	3	5	may	(no response)
22. ever	4	3	even, (no response)	(no response)
23. It	0	1	-	-
24. so	0	4	-	to
25. that	3	3	then	(no response)

Note: C2S2, N=6; MC2S2, N=8.

Table 4.45

Comparison of Errors for C2S3 and MC2S3

Item	Errors		Common Error	
	C2S3	MC2S3	C2S3	MC2S3
1. enough	0	0	-	-
2. to	0	0	-	-
3. for	0	1	-	-
4. days	0	0	-	-
5. any	4	4	all	-
6. time	1	6	-	(no response)
7. the	0	2	-	one
8. has	1	2	-	was
9. enough	0	0	-	-
10. stored	4	5	-	(no response)
11. the	2	0	-	-
12. to	1	1	-	-
13. for	1	0	-	-
14. or	1	0	-	-
15. minutes	1	2	-	(no response)
16. it	0	0	-	-
17. not	1	2	-	-
18. for	0	0	-	-
19. to	0	2	-	-
20. the	1	0	-	-
21. we	0	1	-	-
22. As	3	5	At	Is
23. rule	2	6	-	male, (no response)
24. need	0	2	-	(no response)
25. to	0	1	-	-

Note: C2S3, N=6; MC2S3, N=8.

Table 4.46

Comparison of Errors for C2S4 and MC2S4

Item	Errors		Common Error	
	C2S4	MC2S4	C2S4	MC2S4
1. for	1	1	-	-
2. is	3	0	if	-
3. each	1	1	-	-
4. often	2	6	-	(no response)
5. in	3	2	-	on
6. Another	1	5	-	Further, (no response)
7. Is	4	0	if	-
8. each	1	1	-	-
9. grew	4	2	grow	-
10. separately	2	6	-	(no response)
11. different	3	5	-	(no response)
12. Gradually	2	8	-	Generally, (no response)
13. many	1	3	-	(no response)
14. some	1	4	-	-
15. came	3	5	-	(no response)
16. contact	4	5	(no response)	(no response)
17. other	1	3	-	(no response)
18. When	3	5	What	(no response)
19. happened	1	6	-	(no response)
20. languages	1	5	-	(no response)
21. more	1	5	-	(no response)
22. more	1	5	-	(no response)
23. In	2	3	-	(no response)
24. cases	3	6	(no response)	(no response)
25. a	1	3	-	(no response)

Note: C2S4, N=6; MC2S4, N=8.

surprising since, as has been shown in Table 4.38, 57.4% of all the errors in the MC-Test format are of this type. The reason that there are more errors of this type in the MC-Test format than in the C-Test format is due to the generally higher difficulty of the MC-Test items than the C-Test items. Naturally, many students may have chosen not to guess when they felt the items were too difficult for them.

The errors for the C-Test items and the errors for the MC-Test items are usually different. This is because the responses for the two different types of items are constrained by different orthographic information. For example, the common substitution errors for the word has (#22) were how in C1S2 and was in MC1S2; the common substitution errors for the word one(#6) were off in C1S3 and the in MC1S3; and the common substitution errors for the word that (#12) were then in C2S2 and what in MC2S2.

4.4.2 SUMMARY OF ERROR ANALYSIS

A close examination of the data reveals that in addition to the No Response errors, the other common errors can be further classified into four categories. A brief explanation for each category along with examples from the data are as follows:

1. Common Function Words. This category of error is the most common one for this set of data. When the students come to words containing two to four letters, they will tend to find a word (usually a common function word) that fits the blank by relying heavily on the orthographic information available and with or without much consideration for the syntactic or semantic information present in the text.

   Example:

how	for	her (#5, C1S1)
how	for	now (#25, MC1S1)
that	for	them (#10, C1S1)

what	for	that (#21, MC1S2; #12, MC2S2)
what	for	when (#18, C2S4)
Is	for	It (#9, C1S2)
Is	for	As (#22, MC2S3)

There are two special cases that should be included in this category: itis for this (#19, MC1S4) and tome for some (#7, MC2S1). Although these two errors as "words" appear to be unintelligible, it is likely that the students thought of them as "phrases": it is and to me, respectively. Since these phrases are composed of common function words, it may be appropriate to include them in this category.

2. Words Appearing Unmutilated Elsewhere. This category of errors is found only in MC1S2 and MC2S2. These errors are:

aware	for	there (#3, MC1S2)
there	for	aware (#14, MC2S2)
room	for	from (#6, MC1S2)
see	for	the (#1, #8, & #10, MC1S2)
see	for	are (#13, MC2S2)

The first three common errors are the words that appear unmutilated elsewhere in the same test passage. [24] The word see is included in this category because it appears unmutilated in Subtest 1. Since the topic of Subtest 2 is "Light", the students may have recalled the word see, which is related to the concept of light.

3. More Common Content Words. The errors of this category are found more in the MC-Test format than in the C-Test format. In fact, there is only one error

---

[24] Note that the items in MC1S2 are the unmutilated words in MC2S2, and vice versa. As a result, because the word there is an item and the word aware appears unmutilated in MC1S2, the reverse is true in MC2S2.

of this category found in the C-Test format of the present data which is obviously due to the constraint of the orthographic information present in each type of item. These errors are:

say	for	try (#18, MC1S1)
different	for	difficult (#18, C1S3)
person	for	reason (#1 & #7, MC1S4)
countries	for	centuries (#14, MC1S4)
Generally	for	Gradually (#12, MC2S4)
show	for	know (#4, MC2S1)
male	for	rule (#23, MC2S3)

The first five common errors are all words with higher frequencies than the expected responses.[25] The last two common errors, though not more frequent according to the Rank List, seem to be familiar words for this group of students. The word show is likely to be one of the most common words used in ESL classes. As well, the word male is a word that all adult ESL learners have acquired from the experience in completing application forms for various situations.

4. Wrong Forms of Verbs. Only two common errors of this category are found. Both of them are in the C-Test format. They are:

lives	for	lived (#5, C1S4)
grow	for	grew (#21, C1S4; #9, C2S4)

---

[25] As determined by *The American Heritage Word Frequency Book*'s (Carroll et al., 1971) Rank List.

4.4.3 INTERVIEWS

In order to determine what reading strategies L2 students employed in processing and completing the C-Test/MC-Test, an interview technique[26] was used. Of the 104 L2 students, 28 whose test responses have been error-analyzed, were individually interviewed. The main purpose of the interview was to find out what types of readers these students were and how their reading strategies were related to the errors they made.

The format of the interview was adapted from Devine (1983) who followed Burke (1978). In accordance with the language units they consider important to effective reading, the students can be classified into three major types of readers: sound-, word-, and meaning-oriented. As Devine (1983) points out, these classifications provide an easy way of focusing on readers' theoretical orientations based solely on measurable textual features (p. 97).

Five major questions were designed in order to uncover the students' general attitudes about this type of test, what they thought it measured, what strategies they used, and what types of readers they were. Examples of the questions are shown in Appendix F.

The interviews were audiotaped for re-examination. However, word-for-word transcripts of the interviews were not made, because only the information from these students as groups, rather than as individuals, was needed. The findings are summarized in Table 4.47.

4.4.3.1 ATTITUDE TOWARD THE TEST

The responses to Question #1 demonstrate that, of the 28 interviewees, 21 (75.0 %) have a positive attitude toward the C-Test/MC-Test. When asked why they liked it, the common answer was "It's a good test, and it's easy." This seems to explain why almost all of the students in the LPG liked it, as they felt that they could do the C-Test/MC-Test much better than the Michigan Test. In contrast, four students in the

---

[26] Evidence from Devine (1983) reveals that her ESL subjects at as low level as beginning/low intermediate can "articulate their theoretical orientations towards reading unambiguously enough to be identified as sound-, word-, or meaning-centered" (p. 106).

Table 4.47

Summary of Interview Data

Questions & Answers	HPG	LPG	Total
1. Do you like this type of test?			
a. Yes.	9	12	21
b. Not sure.	4	1	5
c. No.	1	1	2
2. What do you think it tests?			
a. General English ability.	7	8	15
b. Vocabulary.	5	2	7
c. Not sure.	2	4	6
3. What was your strategy in doing the test?			
a. Read word by word.	4	4	8
b. Read the whole sentence first.	4	6	10
c. Read the whole text first.	6	4	10
4. Do you translate when you read in English?			
a. Usually.	4	10	14
b. Sometimes.	4	3	7
c. Rarely.	6	1	7
5. What do you think is most important when you read?			
a. To pronounce the words correctly.	0	4	4
b. To know the meaning of the words.	8	6	14
c. To understand the meaning of the text.	6	4	10

Note: HPG, n=14; LPG, n=14; Total=28.

HPG who found the test not challenging enough (i.e., too easy for them) tended to be more reserved by responding to the question as "I'm not sure." Only two students, one from each group, said that they did not like it, without giving any reason.

4.4.3.2 FACE VALIDITY

Seven students considered the test a vocabulary test, whereas 15 students believed that the test measured not only the knowledge of words but general language ability as well. There were more students in the HPG than in the LGP who thought of the test as a vocabulary test. This finding may imply that because the better readers identify words faster than the poorer readers (Smith, 1971), the former would generally make full use of the orthographic knowledge first, before looking for additional cues from other information sources. The remaining six students, four from the LPG and two from the HPG, simply answered that they were not sure what the test measured. The fact that there are more students from the LPG than from the HPG who could not decide what the test measured seems to imply that high proficiency students have more language awareness than low proficiency students.

Overall, the finding is encouraging because more than half of the students responded that they thought it was a good test of general language abilty, suggesting that the C-Test/MC-Test has fairly good face validity.[27]

4.4.3.3 READING STRATEGIES

Of the 28 students interviewed, eight responded that they read word by word, whereas 10 read the whole sentence and the remaining 10 claimed that they read the whole text first. Between the two proficiency groups, there does not seem to be much difference in the proportions of students using the three strategies. While four students from each group conceded that they attempted the test word by word, four in the HPG and six in the LPG used a sentence context approach, and six in the HPG and four in the LPG tried to apply the meaning of the whole passage as they completed each item. The finding, however, appears to conform to the interactive-compensatory conception, namely, that students in both proficiency levels will use, among other strategies, sentence or whole passage context to find an answer (Stanovich, 1980, p.36). It is not necessarily true that the low proficiency students will always rely solely on the orthographic input (lower-level process), that is, to focus only on each individual word.

---

[27] Hatch & Farhady (1982) conclude that "many people reject the cloze procedure for face validity reasons; they can't accept it as a valid measure of language proficiency" (p. 252). They suggest that: "If we believe that 'on the face of it', the test seems right and that we can defend it as a good test, it has face validity" (p. 252).

The data regarding translation into L1 when reading English (L2) show that half of the 28 students in this study usually translate what they read into their native language. Low proficiency students demonstrate a markedly greater trend toward translation than high proficiency students. The results, therefore, support the assumption that the more proficient the students are in L2 the less they have to translate it into L1 before they can comprehend it.

On the basis of the students' responses to Question #5, what they think is most important for reading, four students can be classified as sound-centered, 14 as word-centered, and the remaining 10 as meaning-centered. The four sound-centered readers are all students in the LPG whose goal for reading was to improve their speaking ability; as a result, for them, pronunciation was central to reading. Although a few students indicated that they consider more than one aspect as important for reading, the elaboration of their answers usually made it clear what type of reading was most predominant. For example, a student in the HPG, designated as a meaning-centered reader, responded that pronunciation is important in reading; however, when reading she wanted to understand the meaning of the text rather than pronounce the words correctly. On the other hand, a word-centered reader, inspite of referring to both pronunciation and meaning as essential for reading, emphasized that vocabulary or word meaning was most important.

It is necessary to note the difference between the two distinguished tasks asked in Questions #3 and #5. In doing the C-Test/MC-Test the students' goal is to complete the missing part of the words, whereas in normal reading their goal may vary according to the purpose of reading. In fact, doing the test is a reading task with its own objective. Therefore, the students' strategies for these two tasks must be different. The results of the interviews indicate that the three types of readers used different strategies in doing the test.

4.4.4 SUMMARY OF INTERVIEWS

As discussed in the preceding section, the proportions of students using the three strategies (in doing the test) are quite similar for the HPG and the LPG; however, when consideration is given to reader type, the distinction between the two proficiency groups becomes more clear. Whereas all 14 students in the HPG are either word-centered or meaning-centered readers, four of the 14 students in the LPG are sound-centered readers. The results of error analysis indicate that the errors made by the HPG are closer to the intended responses than those made by the LPG regarding both the similarity of grammatical function and the retention of meaning. This implies that the high proficiency students who are generally concerned with either word meaning or meaning of the text would try to make more use of both syntactic/structural and semantic information available in the text than the low proficiency students.

Chapter 5

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

The major purpose of this study was to establish the reliability and validity of the C-Test and the MC-Test, which are modified versions of the cloze test. The study also attempted to verify whether different starting points of mutilations would affect the difficulty, reliability, and validity of the tests and to explore the strategies which L2 learners use in restoring the C-Test/MC-Test passages.

In this chapter, the findings are presented in summary form and general conclusions are drawn. A discussion of certain limitations is followed by the possible implications and suggestions for further research.

5.1 SUMMARY OF FINDINGS

The main findings of the investigation are presented under the four headings: Comparability, Reliability, Validity, and Error Analysis and Interviews.

5.1.1 COMPARABILITY

1. Means and item difficulty indices indicate that generally the MC-Test was more difficult than the C-Test. The findings were consistent for both L1 and L2 samples.

2. Since there was no significant difference between the mean scores of Form 1 and Form 2 in either format for both L1 and L2 data, different starting points of mutilations did not affect the difficulty of the tests.

3. Cross-comparison of C-Test and MC-Test item discrimination indices suggests that generally the two item types did not relate to the total test scores in the same way. MC-Test 1 appeared to have more items with high discrimination power ($r > .50$) than C-Test 1 for both L1 and L2 data. MC-Test 2, on the other hand, was superior to C-Test 2 for only the L1 data.

5.1.2 RELIABILITY

1. Coefficients of the Hoyt estimate of reliability indicate that the MC-Test was more reliable than the C-Test.

2. Cronbach's alpha coefficients suggest that in general the degree to which the subtests are homogeneous was satisfactory for all four forms. However, the contradictory results concerning the superiority between the two test formats found in the L1 and L2 data could be considered as an effect of sampling variation.

5.1.3 VALIDITY

1. Results of the factor analyses suggest that the C-Test and the MC-Test had different factor structures. Only 53.6% of the items compared appeared to load in a similar way. Five major characteristics were found to be distinct. These were: adjacency effect, closedness of items, openness of the items, obvious text clues, and practice effect.

2. Adjacency effect was the most common characteristic and emerged more frequently in the MC-Test than in the C-Test. The second most distinctive characteristic was item closedness; namely, the items to which responses were limited due to word type and/or context. This was followed by obvious text clues and practice effect (which could be considered as special cases of item closedness) and finally item openness.

3. Correlations between the C-Test/MC-Test and the Michigan Test, ranging from .423 to .606, can be considered moderately high. This suggests that there was a certain relationship between the C-Test/MC-Test scores and the Michigan Test scores. The C-Test appeared to be a better predictor of performance on the Michigan Test than the MC-Test.

5.1.4 ERROR ANALYSIS AND INTERVIEWS

1. Findings from error analysis show that the percentage of errors which had the same grammatical function as the intended answers was higher in the MC-Test than in the C-Test. On the other hand, the percentage of errors which retained the overall intended meaning was higher in the C-Test than in the MC-Test. Obviously, this is due to the

different orthographic cues present in the two item types.

2. Regarding both similarity of grammatical function and retention of overall meaning, errors made by the HPG were shown to be closer to the intended answers than errors made by the LPG.

3. Common substitution errors can be classified into four categories according to the nature of errors: common function words, words appearing unmutilated elsewhere, more common content words, and wrong forms of verbs.

4. Interviews with 14 subjects reveal that the HPG were either word-centered or meaning-centered readers, whereas four of the 14 subjects in the LPG were found to be sound-centered readers.

5. Proportions of students using the three strategies in restoring the test passages (i.e., read word by word, read the whole sentence first, or read the whole passage first) were quite similar for both proficiency groups.

5.2 CONCLUSIONS AND DISCUSSION

Evidence from all of the analyses demonstrates that the C-Test and the MC-Test were essentially different types of tests. Although both of them appeared to be reliable, valid, and usable (in terms of proficiency level) for both L1 and L2 samples in this study, they were functionally and structurally different.

The MC-Test was more difficult and discriminated better than the C-Test in both L1 and L2 data, concurring with Weaver's (1980) conclusion that "the beginning of words are less predictable than the ends, and therefore more necessary" (p. 49). However, no significant difference in the mean scores was found between Form 1 and Form 2 (where the difference was the starting points of mutilations). This finding supports the recommendation for the C-Test that "the deletions should affect a representative sample of the text" (Klein-Braley & Raatz, 1984, p. 136). Cross-tabulations of item discrimination reveal that C-Test items and MC-Test items did not relate to the total test scores in the same way. Consequently, they were functionally different items; that is, they were not tapping the same abilities.

Reliability coefficients as demonstrated by the Hoyt estimate of reliability favored the MC-Test, illustrating that the MC-Test items were more homogeneous than the C-Test items. This could partly be due to a function of difficulty. In addition, the item discrimination statistics support this finding, since the MC-Test discriminated better than the C-Test. Cronbach's alpha coefficients were also sufficiently high for all four forms, suggesting that the subtests in each test form were satisfactorily homogeneous. This is in agreement with Klein-Braley & Raatz (1984) who reported that in most of the C-Test investigations using Cronbach's alpha, high reliability coefficients (.80 or higher) were obtained (pp. 138-139).

Factor analysis results reveal that the C-Test and the MC-Test had different factor structures. There were more adjacency items (the most common characteristic of items) in the MC-Test than in the C-Test. This interesting finding seems to imply that the restoration of the MC-Test passages requires more reading processes than the restoration of the C-Test passages. In other words, success in processing the MC-Test items would be more dependent on the context with successfully completed items than on the orthographic cues present in the items alone. As claimed by Goodman, Burke, & Sherman (1980), three significant strategies -- predicting, confirming, and integrating -- are usually used by all readers (p. 3). It would be reasonable to hypothesize that the readers would tend to be more involved with these three strategies in processing the MC-Test than the C-Test. While merely the orthographic cues present in the C-Test items may be sufficient, more of the syntactic and semantic information would seem to be necessary in order to confirm the predictions. In other words, in processing the MC-Test items, the readers would tend to have to monitor their hypotheses by integrating all of the information available to them.

A finding that there were more items relating to obvious text clues in the C-Test than in the MC-Test, supports the above discussion. Since an item is normally easier to identify when the first part is given than when the second part is given (Weaver, 1980, p. 49), the readers would likely make a correct prediction for the C-Test item when the word also appeared unmutilated elsewhere in the test passage. On the other hand, in the MC-Test case, the readers would seem to have to gather more information before they could correctly predict, and yet they may not always be successful. An examination of the data reveals that there were more items relating to practice effect in the MC-Test than in the C-Test. Again, item difficulty suggests that when the respondents come to a repeated item, they would tend to monitor their predictions by checking the answer of the previously completed item more in the MC-Test than in the C-Test. The other two characteristics -- item closedness and item openness (which were both more distinct in the C-Test than in the MC-Test) -- also reflect a function of difficulty. When the first part of the word was given, these two opposing characteristics would be more salient than when the second part was given. Thus, these items would tend to cluster according to their characteristic of closedness or openness.

In general, factor structures of the tests reflect that the MC-Test would require the respondents to make more use of other information (i.e., completed-item context, sentence context, or whole passage context) in addition to the orthographic cues present in the item than the C-Test would. According to Goodman, Burke, & Sherman's (1980) model, the readers will generally "select the most significant graphophonic, syntactic, and semantic cues and predict what they believe subsequent graphophonic, syntactic, and semantic structures are going to be" (p. 6). Once predictions are made, the readers would test their hypotheses to see if they are meaningful in order to confirm or disconfirm their predictions. To do this, the readers may have to regress, reread, and pick up additional cues or keep reading in order to build up additional context. On the basis of this model of reading process, it is clear that the processing of the MC-Test passages tend to be more in accordance with the process of normal reading than the processing of the C-Test passages. Raatz's (1984) study indicated that the German L1 C-Test correlated most highly with the orthography test, the grades for German and grammar, and the subtests measuring the ability to structure written linguistic material (p. 133). His finding implies that orthographic knowledge was probably the best predictor of performance on the German L1 C-Test, and vice versa. Therefore, in reference to factor structures, it is reasonable to argue that the MC-Test is a better measure of overall language ability than the C-Test.

Results of the correlational analyses show that the Michigan Test correlated moderately high with the C-Test/MC-Test. The C-Test, however, appeared to predict performance on the Michigan Test better than did the MC-Test. As discussed in Chapter 4 (section 4.3.4), it was expected that the MC-Test would correlate better with the Michigan Test than would the C-Test. This surprising finding could be due to sampling variation since these correlations were calculated from a small number of data points and different supposedly-equated groups.

The common substitution errors found in the C-Test and the MC-Test were different due to the orthographic constraints of the test items (i.e., presence of first half vs. second half). The MC-Test errors were closer to the intended answers than the C-Test errors in the similarity of the grammatical function, but not the retention of the overall intended meaning. As could be expected, the errors made by the HPG were generally closer to the intended answers than the errors made by the LPG. Interview results reveal that, while the HPG were either word-centered or meaning-centered readers, some of the LPG were sound-centered readers. Of the total of 28 students (HPG and LPG), four were designated sound-centered, fourteen word-centered, and the remaining ten meaning-centered. Although the subjects in the present study were relatively more advanced than those in her study, this finding tends to agree with Devine (1983). She reported that of the 19 beginning/lower intermediate level students, six were classified as sound-centered, seven as word-centered, and the other six as meaning-centered (p. 101). Interview data also indicate that the proportion of interviewees using different strategies in restoring the test passages was similar for both the HPG and the LPG. This finding appears to concur with the interactive-compensatory view which suggests that all readers would tend to rely more on other knowledge sources when one particular strategy is inadequate (Stanovich, 1980, p. 36). The results, however, are contrary to the findings of Cziko (1980) who concluded that native speakers and advanced learners of French appeared to use an interactive strategy, whereas intermediate learners would tend to use a bottom-up strategy, relying more on graphic information (p. 113).

In conclusion, the results of this study suggest that the newly-developed MC-Test is empirically superior to the conventional C-Test in several respects. Specifically, the MC-Test discriminates better, has a greater reliability, and is more valid in reference to factor structures. Regarding factor structures, it can be argued that the MC-Test tends to require more of the normal reading processes than does the C-Test. While the restoration of the C-Test passages appears to rely primarily on the orthographic cues available in the items, the processing of the MC-Test items seems to necessitate the use of all the strategies which readers use in the process of normal reading.

5.3 IMPLICATIONS

5.3.1 LIMITATIONS OF THE STUDY

Although a certain amount of research has been conducted on the C-Test, the newly proposed MC-Test is only at its initial stage of development. Given the exploratory nature of the investigation, the results should be taken as suggestive rather than conclusive. Furthermore, the following limitations must be borne in mind in interpreting the results.

1. Since four test forms were used, the number of subjects taking each form was an average of 97 for the L1 sample and 26 for the L2 sample. As a result, factor analysis could not be performed on the L2 data due to an insufficient number of subjects.

2. Due to the design of the study, each subject was assigned to complete only one test form. Thus, a comparison made between each pair of the tests was essentially based on the data from different supposedly-equated groups.

3. Since many L2 subjects who completed the C-Test/MC-Test did not write the Michigan Test, the correlations between them were computed on the data from only 19 to 23 subjects.[21]

---

[21] C-Test 1/Michigan, N=23; MC-Test 1/Michigan, N=21; C-Test 2/Michigan, N=19; and MC-Test 2/Michigan, N=20.

4. Error analysis and interview results must be interpreted with a special caution. As the interview task is elaborate and time-consuming, only 28 L2 subjects, 14 each from the HPG and LPG were used. To correspond with the interview data, error analysis was conducted only on the errors made by these interviewees.

5.3.2 IMPLICATIONS FOR EDUCATION

In their first report on the development of the C-Test, Raatz & Klein-Braley (1981) suggested the following possible uses of the C-Test:

> 1. It could be used at the end of the school year to check on teacher judgments in L1 at the primary level and in foreign language teaching;
> 2. A teacher taking over an unknown class or group could use a C-Test to give him a global idea of the general level of achievement and of the ranking of individual;
> 3. It can be used for selection in cases where selection is necessary, although such procedures are always accompanied by both practical and ethical problems;
> 4. We were able to show in two experiments that C-Tests could be useful as placement tests. A number of investigations into this question are already in progress;
> 5. C-Tests could be used for the diagnosis of learning disturbances in both L1 and L2. Investigations into the use of C-Tests for the diagnosis of dyslexia are being carried out in Duisburg;
> 6. It is difficult to develop tests suitable for the measurement of the level of attainment in German for the children of immigrant workers in the Federal Republic of Germany. A pilot experiment using two classes of Turkish pupils in German secondary schools has produced promising results. (p. 135)

Since the findings of the present study indicate that the MC-Test is superior to the C-Test, it would seem that these suggestions should apply even more to the MC-Test.

Even though the C-Test and the MC-Test (as the names imply) are normally perceived as tests, they can be useful as language exercises. The MC-Test could be beneficial for the teaching of predicting, confirming, and integrating strategies. This area of application of the MC-Test should be brought to the attention of both L1 and L2 language teachers. Unlike the applications for the field of testing, this use of the MC-Test may not require extensive and elaborate investigations before practical application.

5.3.3 SUGGESTIONS FOR FURTHER RESEARCH

The C-Test and the MC-Test are relatively new instruments in language testing. Although the results so far have been promising, further investigations into these tests and especially the MC-Test will be required before any decisive conclusion can be made. The following are some suggestions for further research.

It will be important to replicate this experiment with some modifications. In order that factor analysis can be equally performed, a minimum of 400 subjects each from L1 and L2 samples are required, so that each test form is completed by at least 100 subjects from each sample. It is also necessary to experiment with L1 and L2 samples of a more homogeneous characteristic.[39] For example, a L1 sample may be all from twelfth grade students and a L2 sample from EFL learners who have the same L1 background. The reading passages should be carefully chosen so that the L2 mean score reaches the desirable 50 per cent criterion. The TOEFL, an international standardized proficiency test, should be used as a criterion measure for the L2 sample so that the results can be compared with those obtained from the Michigan Test. As well, the interview should be further developed to focus more profoundly on the differences in processing the C-Test and the MC-Test passages. If these suggested modifications are taken into consideration, the results of a replication would add even further insight into the interpretation of the present study.

In order to further verify the construct validity of the C-Test/MC-Test, these two tests as well as others such as TOEFL and a reading battery could be administered to a consistent group of subjects. After performing factor analysis on the C-Test/MC-Test and the subtests of TOEFL and the reading battery, the loading potential of each test would be revealed.

In addition, further investigation with non-inflectional and/or non-alphabetic languages (e.g., Thai, Chinese) could be valuable in order to determine the extent of applicability and the limitations of the two tests discussed in this study.

---

[39] Raatz & Klein-Braley (1981) note: "It is always possible that results obtained in an investigation are dependent rather on certain characteristics of the samples involved than on the factors interesting the investigator" (p. 133).

BIBLIOGRAPHY

Alderson, J.C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly, 13*(2), 219-227.

Alderson, J.C. (1983). The cloze procedure and proficiency in English as a foreign language. In J.W. Oller, Jr. (Ed.), *Issues in language testing research*. Rowley, Massachusetts: Newbury House, pp. 205-217.

Anderson, J. (1976). *Psycholinguistic experiments in foreign language testing*. Queensland: University of Queensland Press.

Aulls, M. (1981). Developmental considerations for reading research: applications to good and poor reader research. In M.L. Kamil, & M.M. Boswick (Eds.), *Directions in reading: research and instruction*. Thirtieth Yearbook of National Reading Conference. Milwaukee, Wisconsin: N.R.C.

Bachman, L.F. (1982). The trait structure of cloze test scores. *TESOL Quarterly, 16*(1), 61-70.

Bachman, L.F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly, 19*(3), 535-56.

Bachman, L.F., & Palmer, A.S. (1981). Basic concerns in test validation. In J.A.S. Read (Ed.), *Directions in language testing*. Singapore: SEAMEO Regional Language Centre, pp. 41-57.

Berman, R.A. (1984). Syntactic components of the foreign language reading process. In J.C. Alderson & A.H. Urquhart (Eds.), *Reading in a foreign language*. London: Longman, pp. 139-156.

Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly, 20*(3), 463-494.

Boonsathorn, S. (1977). The relationship between understanding English grammatical conjunctions and reading comprehension. Unpublished master's thesis. Chulalongkorn University, Thailand.

Bormuth, J.R. (1963). Cloze as a measure of readability. Proceedings of the International Reading Association 8, pp. 131-134.

Bormuth, J.R. (1965). Optimum sample size and cloze test length in readability measurement. *Journal of Educational Measurement, 2*(1), 111-116.

Brown, J.D.(1983). A closer look at cloze: validity and reliability. In J.W. Oller, Jr. (Ed.), *Issues in language testing research*. Rowley, Massachusetts: Newbury House, pp. 237-250.

Brown, J.D. (1984). A cloze is a cloze is a cloze? In J. Handscombe, R.A. Orem, & B.P. Taylor (Eds.), *On TESOL '83. The question of control*. Washington, D.C.: TESOL, pp. 109-119.

Carroll, J.B. (1971). Language and cognition: current perspectives from linguistics and psychology. In J.F. Laffey & R. Shuy (Eds.), *Language differences: Do they interfere?* Newark, Del.: IRA.

Carroll, J.B. (1972). Defining language comprehension: some speculations. In J.B. Carroll & R.O. Feedle (Eds.), *Language comprehension and the acquisition of knowledge*. New York: Halsted.

Carroll, J.B., Caron, A.S., & Wilds, C.P. (1959). An investigation of cloze items in the measurement of achievement in foreign languages. Report on research conducted under a grant from the College of Education Entrance Examination Board. Cambridge, Massachusette: Graduate School of Education.

Carroll, J.B., Davies, P., & Richman, B. (1971). *The American heritage word frequency book*. New York: American Heritage Publishing Co., Inc.

Carver, R.P. (1975-76). Measuring prose difficulty using the reading scale. *Reading Research Quarterly, 11*(4), 660-685.

Chavez-Oller, M.A., Chihara, T., Weaver, K.A., & Oller, J.W., Jr. (1985). When are cloze items sensitive to constraints across sentences? *Language Learning, 35*,(2), 181-206.

Chihara, T., Oller, J.W., Jr., Weaver, K., & Chavez-Oller, M.A. (1977). Are cloze items sensitive to constraints across sentences? *Language Learning* , *27*, 63-73.

Clark, H.H., & Clark, E.V. (1977). *Psychology and language: an introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich, Inc.

Clark, M. (1979). Reading in Spanish and English: evidence from adult ESL students. *Language Learning, 29*, 121-145.

Coady, J. (1979). A psycholinguistic model of the ESL reader. In R. Mackay, B. Barkman, & R.R. Jordan (Eds.), *Reading in a second language: hypotheses, organization, and practice*. Rowley, Massachusetts: Newbury House, pp. 5-12.

Cohen, A.D., Segal, M., &Weiss Bar-Siman-Tov, R. (1984). The C-Test in Hebrew. *Language Testing, 1*(2), 221-225.

Cooper, C.R., & Petrosky, A.R. (1976). A psycholinguistic view of the fluent reading process. *Journal of Reading*, December, 184-207.

Cooper, M. (1984). Linguistic competence of practised and unpractised non-native readers of English. In J.C. Alderson & A.H. Urquhart (Eds.), *Reading in a foreign language*. London: Longman, pp. 122-135.

Cranney, A.C. (1972-73). The construction of two types of cloze reading tests for college students. *Journal of Reading Behavior, 5*, 60-64.

Cronbach, L.J., & Meehl, P.L. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Cronin, M. (1982). *Contruct validity and the assessment of reading processes*. Unpublished doctoral dissertation, University of Alberta, Alberta.

Cziko, G.A. (1980). Language competence and reading strategies: a comparison of first- and second-language oral reading errors. *Language Learning, 30(1)*, 101-116.

Darnell, D.K. (1968). *The development of an English language proficiency test of foreign students using a clozentropy procedure* (Final report to DHEW Bureau No. BP-7-H-010). Boulder, Colorado: University of Colorado. (ERIC ED 024 039)

Devine, J. (1983). ESL readers' internalized models of the reading process. In J. Handscombe, R.A. Orem, & B.P. Taylor (Eds.), *On TESOL '83: the question of control*. Washington, D.C.: TESOL.

Foley, J.A. (1983). More questions on assumptions about cloze testing. *RELC Journal, 14*(1), 57-69.

Freeland, J. (1979). Text type as a factor in the cloze testing of foreign languages. *BAAL Newsletter, No. 8*.

Fromkin, V. & Rodman, R. (1983). *An introduction to language*. New York: Holt, Rinehart and Winston.

Fry, E. (1964). *Reading faster: a drill book*. Cambridge: Cambridge University Press.

Fry, E. (1977). *Graph for estimating readability-extended*. New Brunswick, New Jersey: Rutgers University Reading Center.

Goodman, K.S. (1967). Reading a psycholinguistic guessing game. In F.V. Gollasch (Ed.), *Language and literacy: the selected writings of Kenneth Goodman. Volume I*. London: Routledge & Kegan Paul, pp. 33-43.

Goodman, K., & Niles, O. (1970). Behind the eye: what happens in reading. In F.V. Gollasch (Ed.), *Language and literacy: the selected writings of Kenneth Goodman. Volume II*. London: Routledge & Kegan Paul, pp. 99-124.

Goodman, Y.M., & Burke, C.L. (1972). *Reading miscue inventory manual: procedure for diagnosis and evaluation*. New York: Macmillan Publishing Co., Inc.

Goodman, Y.M., Burke, C., & Sherman, B. (1980). *Reading strategies: focus on comprehension*. New York: Holt, Rinehart and Winston.

Gradman, H.S., & Spolsky, B. (1975). Reduced reducdancy testing: a progress report. In R.L. Jones, & B. Spolsky (Eds.), *Testing language proficiency*. Arlington, Va.: Center for Applied Linguistics, pp. 59-70.

Greene, F.P. (1965). Modification of the cloze procedure and changes in reading test performances. *Journal of Educational Measurement, 2*(2), 213-217.

Greene, F.P., Cranney, A.G., Bormuth, J., Bloom, R., Rankin, E., Weaver, W., & Brown, E. (1967). Cloze symposium. In G.B. Schick & M.M. May (Eds.), *Multidisciplinary aspects of college-adult reading*. Seventeenth Yearbook of the National Reading Conference. Milwaukee, Wisconsin: N.R.C.

Grove, M.P. (1981). Psycholinguistic theories and ESL reading. In C.W. Twyford, W. Diehl, & K. Feathers (Eds.), *Reading English as a second language: moving from theory. Monograph in language and reading studies, No. 4*. Indiana University School of Education, pp. 3-20.

Grover, C.C., & Anderson, D.G. (1961). *New practice readers, Book G*. U.S.A.: McGraw-Hill.

Grover, C.C., Kinkead, E.B., & Anderson, D.G. (1962). *New practice readers, Book F.* U.S.A.: McGraw-Hill.

Harste, J., & Burke, C. (1977). A new hypothesis for reading teacher education. In P.D. Pearson (Ed.), *Reading: research, theory and practice.* Mason Publishing.

Harste, J.C., Woodward, V.A., & Burke, C.L. (1984). *Language stories & literacy lessons.* Portsmouth, N.H.: Heinemann Educational Books.

Hatch, E., & Farhady, H. (1982). *Research design and statistics for applied linguistics.* Rowley, Massachusetts: Newbury House.

Hill, L.A. (1967). *Elementary stories for reproduction.* London: Oxford University Press.

Hosenfeld, C. (1977). A preliminary investigation of the reading strategies of successful and nonsuccessful second language learners. *System, 5*(2), 111-123.

Johnson, R.K., (1981). Questioning some assumptions about cloze testing. In J.A.S. Read (Ed.), *Directions in language testing.* Singapore: SEAMEO Regional Language Centre, pp. 177-206.

Jonz, J. (1976). Improving on the basic egg: the M-C cloze. *Language Learning, 26*(2), 255-265.

Klare, G.R. (1974-1975). Assessing readability. *Reading Research Quarterly, 10*(1), 62-102.

Klein-Braley, C. (1981). Emprical investigations of cloze tests : an examination of the validity of cloze tests as tests of general language proficiency in English for German university students. Unpublished doctoral dissertation. Duisburg, Germany.

Klein-Braley, C. (1983). A cloze is a cloze is a question. In J.W. Oller, Jr. (Ed.), *Issues in language testing research.* Rowley, Massachusetts: Newbury House, pp. 218-228.

Klein-Braley, C. (1985). A cloze-up on the C-Test: a study in the construct validation of authentic tests. *Language Testing 2*(1), 76-104.

Klein-Braley, C., & Raatz, U. (1984). A survey of research on the C-test. *Language Testing, 1*(2), 134-146.

Kooi, B.Y., Schutz, R.D., & Baker, R.L. (1965). Spelling errors and the serial position effect. *Journal of Educational Psychology, 56,* 334-336.

Krashen, S.D. (1981). *Second language acquisition and second language learning.* Oxford: Pergamon.

Krashen, S.D. (1982). *Principles and practice in second language acquisition.* Oxford: Pergamon.

Lee, Y.P. (1985). Investigating the validity of the cloze score. In Y.P. Lee, A.C.Y.Y. Fok, R. Lord & G. Low (Eds.), *New directions in language testing.* Oxford: Pergamon Press, pp. 137-147.

Levenston, E.A., Nir, R., & Blum-Kulka, S. (1984). Discourse analysis and the testing of reading comprehension by cloze techniques. In A.K. Pugh & J.M. Ulijn (Eds.), *Reading for professional purposes: studies and practices in native and foreign languages.*

London: Heinemann Educational Books, pp. 202-212.

Loevinger, J. (1967). Objective tests as instruments of psychological theory. In D.N. Jackson & S. Messick (Eds.), *Problems in human assessment*. New York: McGraw-Hill, pp. 78-123.

Markham, P.L. (1985). The rational deletion cloze and global comprehension in German. *Language Learning, 35*(3), 423-430.

Miller, G.R., & Coleman, E.B. (1967). A set of thirty-six prose passages calibrated for complexity. *Journal of Verbal Learning and Verbal Behavior, 6*, 851-854.

Mulaik, S.A. (1972). *The foundation of factor analysis.* New York: McGraw-Hill Book Company.

Nelson, L.R. (1974). *Guide to LERTAP use and interpretation.* Department of Education, University of Otago, Dunedin, New Zealand.

Neville, M.H., & Pugh, A.K. (1976-77). Content in reading and listening: variations in approach to cloze tasks. *Reading Research Quarterly, 12*,(1), 13-31.

Ohnmacht, F.W. Weaver, W.W., & Kohler, E.T. (1970). Cloze and closure: a factorial study. *Journal of Psychology, 74*, 205-217.

Oller, J.W., Jr. (1972). Scoring methods and difficulty levels for tests of proficiency in English as a second language. *Modern Language Journal, 56*, 151-158.

Oller, J.W., Jr. (1973a). Cloze tests of second language proficiency and what they measure. *Language Learning, 23*, 105-118.

Oller, J.W., Jr. (1973b). Discrete point tests versus tests of integrative skills. In Oller & Richards (Eds.), *Focus on the learner: pragmatic perspectives for the language teacher.* Rowley, Massachusette: Newbury House, pp. 184-199.

Oller, J.W., Jr. (1975). Cloze, discourse, and approximation to English. In M.K. Burt & H.C. Dulay (Eds.), *New directions in TESOL.* Washington, D.C.: TESOL, pp. 345-356.

Oller, J.W., Jr. (1979). *Language tests at school: a pragmatic approach.* London: Longman.

Oller, J.W., Jr. (1983). Evidence for a general language proficiency factor: an expectancy grammar. In J.W. Oller, Jr. (Ed.), *Issues in language testing research.* Rowley, Massachusetts: Newbury House, pp. 3-10.

Oller, J.W., Jr., Bowen, J.D., Dien, T.T., & Mason,V.W. (1972). Cloze tests in English, Thai, and Vietnamese: native and non-native performance. *Language Learning, 22*, 1-15.

Oller, J.W., Jr., & Conrad, C. (1971). The cloze technique and ESL proficiency. *Language Learning, 21*, 181-195.

Oller, J.W., Jr., & Inal, N. (1975). A cloze test of English prepositions. In L. Palmer & B. Spolsky (Eds.), *Papers on language testing.* Washington, D.C.: TESOL, pp. 38-49.

Osgood, C.E. (1959). The representatiunal model and relevant research methods. In I. de

Sola Pool (Ed.), *Trends in content analysis*. Urbana: University of Illinois Press, pp. 33-88.

Ozete, O. (1977). The cloze procedure: a modification. *Foreign Language Annals, 10*(5), 565-568.

Porter, D. (1976). Modified cloze procedure: a more valid reading comprehension test. *English Language Teaching Journal, 30*(2), 151-155.

Propst, I.K., Jr., & Baldauf, R.B., Jr. (1979). Use matching cloze tests for elementary ESL students. *The Reading Teacher, 32*(6), 683-690.

Propst, I.K., Jr., & Baldauf, R.B., Jr. (1981). A psycholinguistic rationale of measuring beginning ESL reading with matching cloze tests. *RELC Journal, 12*(1), 85-89.

Pugh, A.K., & Blenkhoorn, P. (1984). Computer analysis of cloze responses: a comprehensive system and some findings from its application. Paper presented at the 11th ALLC Conference, Louvain-Neuve.

Quirk, R. & Greenbaum, S. (1973). *A concise grammar of contemporary English.* New York: Harcourt Brace Jovanovich, Inc.

Raatz, U. (1984). The validity of C-Tests. In T. Culhane, C. Klein-Braley, & D.K. Stevenson (Eds.), *Practice and problems in language testing 7*. Department of Language and Linguistics Occasional Papers No. 29. Colchester: University of Essex, pp. 124-139.

Raatz, U. & Klein-Braley, C. (1981). The C-Test -- a modification of the cloze procedure. In T. Culhane, C. Klein-Braley, & D.K. Stevenson (Eds.), *Practce and problems in language testing*. Department of Language and Linguistics Occasional Papers No. 26. Colchester: University of Essex, pp. 113-138.

Ramanauskas, S. (1972). The responsiveness of cloze readability measures to linguistic variables operating over segments of text longer than a sentence. *Reading Research Quarterly, 8* , 79-91.

Rand, E. (1978). The effects of test length and scoring method on the precision of cloze test scores. *Workpapers in teaching English as a second language, 12*, 62-71. Los Angeles: UCLA.

Rankin, E.F., Jr. (1965). Closure and the cloze procedure. Third and Fourth Yearbooks of the North Central Reading Association. Minneapolis: University of Minnesota.

Rankin, E.F., & Thomas, S. (1980). Contextual constraints and the construct validity of the cloze procedure. Twenty-ninth Yearbook of the National Reading Conference.

Rhodes, L. (1979). Comprehension and predictability: an analysis of beginnibg reading materials. In Harste & Carey (Eds.), *New perrspectives on comprehension.* Bloomington: Indiana University School of Education.

Ruddell, R.B., Taylor, M., & Adams, P.J. (1978). *Upside and down*. Boston Rockleigh, N.J.: Allyn and Bacon, Inc.

Ruddle, E.B. (1964). A study of the cloze comprehension technique in relation to structurally controlled reading material. Procedings of the International Reading Association 9, pp. 298-303.

Scholz, G.E., & Scholz, C.M. (1981). Multiple-choice cloze tests of EST discourse: an exploration. Paperpresented at the Fifteenth Annual TESOL Convention. (ERIC Document Reproduction Service No. ED 208 656.

Shanahan, T., Kamil, M.L.; & Tobin, A.W. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly, 17*, 229-255.

Shannon, C.E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Simon, D.P. & Simon, H.A. (1973). Alternative use of phonemic information in spelling. *Review of Educational Research, 43*(1), 115-137.

Singer, H.S., & Ruddell, R.B. (Eds.). (1976). *Theoretical models and processes of reading (2nd ed.)*. Newark, Delaware: International Reading Association.

Singer, H.S., & Ruddell, R.B. (Eds.). (1985). *Theoretical models and processes of reading (3rd ed.)*. Newark, Delaware: International Reading Association.

Smith, E.B., Goodman, K.S., & Meredith, R. (1970). *Language and thtinking in the elementary school.* New York: Holt, Rinehart & Winston.

Smith, F. (1978). *Understanding reading*. Toronto: Holt, Rinehart and Winston.

Stanovich, K.E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly 16*(1), 32-71.

Stansfield, C., & Hansen, J. (1983). Field dependence-independence as a variable in second language cloze test performance. *TESOL Quarterly* , *17*(1), 29-38.

Taylor, W.L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly, 30*, 415-453.

Taylor, W.L. (1956). Recent development in the use of "cloze procedure". *Journalism Quarterly, 33*, 42-48, 99.

Taylor, W.L. (1957). Cloze readability scores as indicies of individual differences in comprehension and aptitude. *Journal of Applied Psychology, 41*, 19-26.

Taylor W.L. (1972). Relative influences of preceding, following, and surrounding contexts on cloze procedure. Twenty-first Yearbook of the National Reading Conference, pp. 66-73.

Taylor, W.L., & Waldman, I.N. (1969). Latency and focus methods of cloze quantification. In G.B. Schick & M.M. May (Eds.), *Reading: process and pedagogy*. Nineteenth Yearbook of the National Reading Conference. Milwaukee, Wisconsin: N.R.C.

Ulijn, J.M. (1984). Reading for professional purposes: psycholinguistic evidence in a cross-linguistic perspective. In A.K. Pugh & J.M. Ulijn (Eds.), *Reading for professional purposes: studies and practices in native and foreign languages.* London: Heinemann Educational Books, pp. 66-81.

Weaver, C. (1980). *Psycholinguistics and reading: from process to practice.* Cambridge, Massachusetts: Winthrop Publishers, Inc.

Weaver, W. (1949). The mathematics of communication. *Scientific American, 181*(1), 11-15.

Weaver, W.W. (1965). Theoretical aspects of the cloze procedure. Fourteenth Yearbook of the National Reading Conference, pp. 115-132.

Willemsen, E.W. (1974). *Understanding statistical reasoning: how to evaluate research literature in the behavioral sciences.* San Francisco: W.H. Freeman and Company.

Youngman, M.B. (1979). *Analysing social and educational research data.* London: McGraw-Hill Book Company (UK) Limited.

APPENDICES

APPENDIX A

TESTS AND GUIDELINES FOR ADMINISTRATION

GUIDELINES FOR TEST ADMINISTRATION

Department of Elementary Education
University of Alberta
Phone: 432-3840

May 21, 1986

Dear Instructor:

Re: *Guidelines for the Michigan Test Administration*

1. Please remind the students to write their names, native languages, and their class level on the ANSWER SHEET. This is very important because the Michigan Test scores will be correlated with the C-Test/MC-Test scores. (The C-Test/MC-Test will be administered to the students on the following day.)

2. Please read the instructions and examples for all three sections with your students before they begin the test. The low level students may need your help in understanding the instructions.

3. The students have only 75 minutes to complete the entire test.

4. Please take every precaution to ensure that each student does his/her own work. Cheating will adversely affect the research project. You may wish to organize the classroom in an individual seating arrangement.

5. If the students wish to know their results, please have them write their names on the attached sheet.

Thank you for administering the test. Your assistance is greatly appreciated.

Sincerely,

Somsak Boonsathorn

Department of Elementary Education
University of Alberta
Phone 432-3840

May 22, 1986

Dear Instructor:

Re: *Guidelines for administering the C-Test/ MC-Test*

1. There are four parallel forms of tests (C1, MC1, C2, & MC2). Please distribute them as arranged so that no two adjacent students in the row have the same form.

2. Have the students complete all the items on the INFORMATION sheet. Some students may need your help in doing this.

3. Please read the test instructions with your students and emphasize that each blank (-) represents one missing letter. Have the students practice the sample tests which follow to ensure that they understand the task before they begin the real tests.

4. There is no time limit. The task may take only half an hour or up to an hour and a half to complete for some students.

5. Please encourage all students to finish all the items. (There are four reading passages, 25 items each.)

6. Please take every precaution to ensure that each student does his/her own work. Cheating will adversely affect the research project. You may wish to organize the classroom in an individual seating arrangement.

Thank you for administering the task. Your assistance is greatly appreciated.

Sincerely,

Somsak Boonsathorn

N.B. Attached are the information sheet, the test instructions sheet, and the student name sheet.

C-TEST 1

(C-TEST 1)

INFORMATION

THIS INFORMATION IS NEEDED FOR RESEARCH PURPOSES ONLY AND WILL BE KEPT STRICTLY CONFIDENTIAL. AS WELL, TEST SCORES WILL NOT BE RELEASED AND ALL TESTEES SHALL REMAIN ANONYMOUS.

INSTRUCTIONS: Please supply the information requested below.

Name: ______________________________ Sex: ____________

ESL class level at Alberta Vocational Centre: ____________

Native country: ______________________

Native language: ______________________

Other language(s) spoken: ______________________

Language spoken at home: ______________________

Number of years of living in Canada: ________

Number of years of learning English in Canada: ______

Number of years of learning English in native country: _______

Number of years of schooling in native country: ________

Certificate(s) or Diploma(s) received: ________________________

C-TEST/MC-TEST

INSTRUCTIONS:

In each of the four passages below, half of every second word has been taken out and an appropriate space has been put in its place. You are to complete the passages by filling in the missing part of each word.

You should:

1. try to fill in every blank; and
2. guess if you are not sure.

There is no time limit but it may take about thirty to fifty minutes to complete these problems.

Try the following examples:

C-TEST

How do we show that we are friends? We sh_ _ we a_ _ friends wh_ _ we sm_ _ _. We sh_ _ we a_ _ friends wh_ _ we he_ _. You c_ _ help _ _ friend. Y_ _ can b_ a fri_ _ _. Friends make the world a very good world for you and for me.

MC-TEST

How do we show that we are friends? We _ _ow we _ _e friends _ _en we _ _ _le. We _ _ow we _ _e friends _ _en we _ _lp. You _ _n help _ friend. _ _u can _e a _ _ _end. Friends make the world a very good world for you and for me.

C-TEST 1

Passage 1 (C1S1)

A woman was having some trouble with her heart, so she went to see the doctor. He w__ a n__ doctor, a__ did n__ know h__, so h_ first as___ some ques_____, and o__ of th__ was, "H__ old a__ you?"

"We__," she answ____, "I do___ remember, doc___, but _ will t__ to th___." She tho____ for _ minute a__ then sa__, "Yes, _ remember n__, doctor! When I married, I was eighteen years old, and my husband was thirty.

C-TEST 1

Passage 2 (C1S2)

If you are in a totally dark room and you turn on a flashlight, it will seem that a circle of light appears instantly on the opposite wall. Actually, t_ _ light i_ not th_ _ _ instantly. I_ has t_ travel fr_ _ your flash_ _ _ _ _ to t_ _ wall. I_ crosses t_ _ entire ro_ _ so fa_ _ that y_ _ are n_ _ aware th_ _ it i_ taking a_ _ time.

Li_ _ _ travels fas_ _ _ than anyt_ _ _ _ else th_ _ man h_ _ ever meas_ _ _ _. It tra_ _ _ _ so fa_ _ that the human eye is unable to perceive it move. Scientists have calculated that light travels at a speed of almost three hundred thousand kilometers per second.

C-TEST 1

Passage 3 (C1S3)

The body of the average adult has in storage enough food to last for several weeks. It h_ _ enough wa_ _ _ to la_ _ for sev_ _ _ _ days. A_ any o_ _ time, how_ _ _ _, the bo_ _ has on_ _ enough oxy_ _ _ stored i_ the lu_ _ _ to la_ _ for th_ _ _ or fo_ _ minutes! Fortu_ _ _ _ _ _ _, it i_ not diff_ _ _ _ _ for u_ to acq_ _ _ _ the oxy_ _ _ we ne_ _. As _ rule, w_ need on_ _ to breathe in the air around us for an adequate supply. The amount of oxygen needed by a person depends upon his activities.

C-TEST 1

Passage 4 (C1S4)

The speech of every group of people has developed differently. One rea___ for th__ is th__ each gr___ often li___ in isol_____. Another rea___ is th__ each lang____ grew u_ separately un___ different condi_____. Gradually, ov__ many cent_____, some gro___ came i_ contact wi__ other gro___. When th__ happened, th___ languages gr__ more a__ more al___. In ot___ cases, how____, a group would break apart, and new languages would grow from the original one. The new languages, though distinct in many ways, would maintain similarities to one another.

MC-TEST 1

(MC-TEST 1)

INFORMATION

THIS INFORMATION IS NEEDED FOR RESEARCH PURPOSES ONLY AND WILL BE KEPT STRICTLY CONFIDENTIAL. AS WELL, TEST SCORES WILL NOT BE RELEASED AND ALL TESTEES SHALL REMAIN ANONYMOUS.

INSTRUCTIONS: Please supply the information requested below.

Name: ______________________ Sex: __________

ESL class level at Alberta Vocational Centre: __________

Native country: ________________

Native language: ________________

Other language(s) spoken: ________________

Language spoken at home: ________________

Number of years of living in Canada: ______

Number of years of learning English in Canada: _____

Number of years of learning English in native country: ______

Number of years of schooling in native country: ______

Certificate(s) or Diploma(s) received: ________________

C-TEST/MC-TEST

INSTRUCTIONS:

In each of the four passages below, half of every second word has been taken out and an appropriate space has been put in its place. You are to complete the passages by filling in the missing part of each word.

You should:

1. try to fill in every blank; and
2. guess if you are not sure.

There is no time limit but it may take about thirty to fifty minutes to complete these problems.

Try the following examples:

C-TEST

How do we show that we are friends? We sh_ _ we a_ _ friends wh_ _ we sm_ _ _. We sh_ _ we a_ _ friends wh_ _ we he_ _. You c_ _ help _ friend. Y_ _ can b_ a fri_ _ _. Friends make the world a very good world for you and for me.

MC-TEST

How do we show that we are friends? We _ _ow we _ _e friends _ _en we _ _ _le. We _ _ow we _ _e friends _ _en we _ _lp. You _ _n help _ friend. _ _u can _e a _ _ _end. Friends make the world a very good world for you and for me.

MC-TEST 1

Passage 1 (MC1S1)

A woman was having some trouble with her heart, so she went to see the doctor. He _ _s a _ _w doctor, _ _d did _ _t know _ _r, so _e first _ _ _ed some _ _ _ _tions, and _ _e of _ _em was, "_ _w old _ _e you?"

"_ _ll," she _ _ _ _ered, "I _ _ _'t remember, _ _ _tor, but _ will _ _y to _ _ _nk." She _ _ _ _ght for _ minute _ _d then _ _id, "Yes, _ remember _ _w, doctor! When I married, I was eighteen years old, and my husband was thirty.

MC-TEST 1

Passage 2 (MC1S2)

If you are in a totally dark room and you turn on a flashlight, it will seem that a circle of light appears instantly on the opposite wall. Actually, _ _e light _s not _ _ _re instantly. _t has _o travel _ _om your _ _ _ _ _light to _ _e wall. _t crosses _ _e entire _ _om so _ _st that _ _u are _ _t aware _ _at it _s taking _ _y time.

_ _ _ht travels _ _ _ter than _ _ _ _hing else _ _at man _ _s ever _ _ _ _ured. It _ _ _ _els so _ _st that the human eye is unable to perceive it move. Scientists have calculated that light travels at a speed of almost three hundred thousand kilometers per second.

MC-TEST 1

Passage 3 (MC1S3)

The body of the average adult has in storage enough food to last for several weeks. It _ _s enough _ _ _er to _ _st for _ _ _ _ral days. _t any _ _e time, _ _ _ _ver, the _ _dy has _ _ly enough _ _ _gen stored _n the _ _ _gs to _ _st for _ _ _ee or _ _ur minutes! _ _ _ _ _ _ately, it _s not _ _ _ _ _cult for _s to _ _ _ _ire the _ _ _gen we _ _ed. As _ _ rule, _e need _ _ly to breathe in the air around us for an adequate supply. The amount of oxygen needed by a person depends upon his activities.

MC-TEST 1

Passage 4 (MC1S4)

The speech of every group of people has developed differently. One _ _ _son for _ _is is _ _at each _ _ _up often _ _ _ed in _ _ _ _ _tion. Another _ _ _son is _ _at each _ _ _ _uage grew _p separately _ _ _er different _ _ _ _ _tions. Gradually, _ _er many _ _ _ _ _ries, some _ _ _ups came _n contact _ _th other _ _ _ups. When _ _is happened, _ _ _ir languages _ _ew more _ _d more _ _ _ke. In _ _ _er cases, _ _ _ _ver, a group would break apart, and new languages would grow from the original one. The new languages, though distinct in many ways, would maintain similarities to one another.

C-TEST 2

(C-TEST 2)

INFORMATION

THIS INFORMATION IS NEEDED FOR RESEARCH PURPOSES ONLY AND WILL BE KEPT STRICTLY CONFIDENTIAL. AS WELL, TEST SCORES WILL NOT BE RELEASED AND ALL TESTEES SHALL REMAIN ANONYMOUS.

INSTRUCTIONS: Please supply the information requested below.

Name: ______________________ Sex: __________

ESL class level at Alberta Vocational Centre: __________

Native country: ________________

Native language: ________________

Other language(s) spoken: ________________

Language spoken at home: ________________

Number of years of living in Canada: ______

Number of years of learning English in Canada: ______

Number of years of learning English in native country: ______

Number of years of schooling in native country: ______

Certificate(s) or Diploma(s) received: __________________

C-TEST/MC-TEST

INSTRUCTIONS:

In each of the four passages below, half of every second word has been taken out and an appropriate space has been put in its place. You are to complete the passages by filling in the missing part of each word.

You should:

1. try to fill in every blank; and
2. guess if you are not sure.

There is no time limit but it may take about thirty to fifty minutes to complete these problems.

Try the following examples:

C-TEST

How do we show that we are friends? We sh_ _ we a_ _ friends wh_ _ we sm_ _ _. We sh_ _ we a_ _ friends wh_ _ we he_ _. You c_ _ help _ friend. Y_ _ can b_ a fri_ _ _. Friends make the world a very good world for you and for me.

MC-TEST

How do we show that we are friends? We _ _ow we _ _e friends _ _en we _ _ _le. We _ _ow we _ _e friends _ _en we _ _lp. You _ _n help _ friend. _ _u can _e a _ _ _end. Friends make the world a very good world for you and for me.

C-TEST 2

Passage 1 (C2S1)

A woman was having some trouble with her heart, so she went to see the doctor. He was _ new doc___, and d__ not kn__ her, s_ he fi___ asked so__ questions, a__ one o_ them w__, "How o__ are y__?"

"Well," s__ answered, "_ don't reme____, doctor, b__ I wi__ try t_ think." S__ thought f__ a min___ and th__ said, "Y__, I reme____ now, doc___! When I married, I was eighteen years old, and my husband was thirty.

C-TEST 2

Passage 2 (C2S2)

If you are in a totally dark room and you turn on a flashlight, it will seem that a circle of light appears instantly on the opposite wall. Actually, the li_ _ _ is n_ _ there inst_ _ _ _ _. It h_ _ to tra_ _ _ from yo_ _ flashlight t_ the wa_ _. It cro_ _ _ _ the ent_ _ _ room s_ fast th_ _ you a_ _ not aw_ _ _ that i_ is tak_ _ _ any ti_ _.

Light tra_ _ _ _ faster th_ _ anything el_ _ that m_ _ has ev_ _ measured. I_ travels s_ fast th_ _ the human eye is unable to perceive it move. Scientists have calculated that light travels at a speed of almost three hundred thousand kilometers per second.

C-TEST 2

Passage 3 (C2S3)

The body of the average adult has in storage enough food to last for several weeks. It has eno_ _ _ water t_ last f_ _ several da_ _. At a_ _ one ti_ _, however, t_ _ body h_ _ only eno_ _ _ oxygen sto_ _ _ in t_ _ lungs t_ last f_ _ three o_ four min_ _ _ _! Fortunately, i_ is n_ _ difficult f_ _ us t_ acquire t_ _ oxygen w_ need. A_ a ru_ _, we ne_ _ only t_ breathe in the air around us for an adequate supply. The amount of oxygen needed by a person depends upon his activities.

C-TEST 2

Passage 4 (C2S4)

The speech of every group of people has developed differently. One reason f_ _ this i_ that ea_ _ group of_ _ _ lived i_ isolation. Ano_ _ _ _ reason i_ that ea_ _ language gr_ _ up separ_ _ _ _ _ under diff_ _ _ _ _ conditions. Grad_ _ _ _ _, over ma_ _ centuries, so_ _ groups ca_ _ in con_ _ _ _ with ot_ _ _ groups. Wh_ _ this happ_ _ _ _, their lang_ _ _ _ _ grew mo_ _ and mo_ _ alike. I_ other ca_ _ _, however, _ group would break apart, and new languages would grow from the original one. The new languages, though distinct in many ways, would maintain similarities to one another.

MC-TEST 2

(MC-TEST 2)

INFORMATION

THIS INFORMATION IS NEEDED FOR RESEARCH PURPOSES ONLY AND WILL BE KEPT STRICTLY CONFIDENTIAL. AS WELL, TEST SCORES WILL NOT BE RELEASED AND ALL TESTEES SHALL REMAIN ANONYMOUS.

INSTRUCTIONS: Please supply the information requested below.

Name: ______________________________ Sex: ____________

ESL class level at Alberta Vocational Centre: ____________

Native country: ______________________

Native language: ______________________

Other language(s) spoken: ______________________

Language spoken at home: ______________________

Number of years of ~~living~~ in Canada: ________

Number of years of learning English in Canada: ______

Number of years of learning English in native country: _______

Number of years of schooling in native country: ________

Certificate(s) or Diploma(s) received: _________________________

C-TEST/MC-TEST

INSTRUCTIONS:

In each of the four passages below, half of every second word has been taken out and an appropriate space has been put in its place. You are to complete the passages by filling in the missing part of each word.

You should:

1. try to fill in every blank; and
2. guess if you are not sure.

There is no time limit but it may take about thirty to fifty minutes to complete these problems.

Try the following examples:

C-TEST

How do we show that we are friends? We sh_ _ we a_ _ friends wh_ _ we sm_ _ _. We sh_ _ we a_ _ friends wh_ _ we he_ _. You c_ _ help _ friend. Y_ _ can b_ a fri_ _ _. Friends make the world a very good world for you and for me.

MC-TEST

How do we show that we are friends? We _ _ow we _ _e friends _ _en we _ _ _le. We _ _ow we _ _e friends _ _en we _ _lp. You _ _n help _ friend. _ _u can _e a _ _ _end. Friends make the world a very good world for you and for me.

MC-TEST 2

Passage 1 (MC2S1)

A woman was having some trouble with her heart, so she went to see the doctor. He was _ new _ _ _tor, and _ _d not _ _ow her, _o he _ _ _st asked _ _me questions, _ _d one _f them _ _s, "How _ _d are _ _u?"

"Well," _ _e answered, "_ don't _ _ _ _mber, doctor, _ _t I _ _ll try _o think." _ _e thought _ _r a _ _ _ute and _ _en said, "_ _s, I _ _ _ _mber now, _ _ _tor! When I married, I was eighteen years old, and my husband was thirty.

MC-TEST 2

Passage 2 (MC2S2)

If you are in a totally dark room and you turn on a flashlight, it will seem that a circle of light appears instantly on the opposite wall. Actually, the _ _ _ht is _ _t there _ _ _ _ _ntly. It _ _s to _ _ _vel from _ _ur flashlight _o the _ _ll. It _ _ _ _ses the _ _ _ire room _o fast _ _at you _ _e not _ _ _re that _t is _ _ _ing any _ _me.

Light _ _ _ _els faster _ _an anything _ _se that _ _n has _ _er measured. _t travels _o fast _ _at the human eye is unable to perceive it move. Scientists have calculated that light travels at a speed of almost three hundred thousand kilometers per second.

MC-TEST 2

Passage 3 (MC2S3)

The body of the average adult has in storage enough food to last for several weeks. It has _ _ _ugh water _o last _ _r several _ _ys. At _ _y one _ _me, however, _ _e body _ _s only _ _ _ugh oxygen _ _ _red in _ _e lungs _o last _ _r three _r four _ _ _ _tes! Fortunately, _t is _ _t difficult _ _r us _o acquire _ _e oxygen _e need. _s a _ _le, we _ _ed only _o breathe in the air around us for an adequate supply. The amount of oxygen needed by a person depends upon his activities.

MC-TEST 2

Passage 4 (MC2S4)

The speech of every group of people has developed differently. One reason _ _r this _s that _ _ch group _ _ _en lived _n isolation. _ _ _ _her reason _s that _ _ch language _ _ew up _ _ _ _ _ately under _ _ _ _ _rent conditions. _ _ _ _ _ally, over _ _ny centuries, _ _me groups _ _me in _ _ _ _act with _ _ _er groups. _ _en this _ _ _ _ened, their _ _ _ _ _ages grew _ _re and _ _re alike. _n other _ _ _es, however, _ group would break apart, and new languages would grow from the original one. The new languages, though distinct in many ways, would maintain similarities to one another.

APPENDIX B

ITEM STATISTICS

Table B-1

Comparison of Item Statistics for C1S1 and MC1S1

Item	L1 Data				L2 Data			
	Difficulty		Discrimination		Difficulty		Discrimination	
	C1S1	MC1S1	C1S1	MC1S1	C1S1	MC1S1	C1S1	MC1S1
1. was	98.0	92.6	0.71	0.63	92.3	84.6	0.52	0.20
2. new	87.3	76.8	0.68	0.55	57.7	53.8	0.88	0.62
3. and	98.0	83.2	0.76	0.39	92.3	84.6	0.62	0.46
4. not	99.0	94.7	0.07	1.07	84.6	73.1	0.35	0.78
5. her	49.0	*89.5	0.35	0.85	46.2	*76.9	0.04	0.70
6. he	100.0	96.8	0.00	0.85	100.0	84.6	0.00	0.43
7. asked	100.0	82.1	0.00	0.69	92.3	69.2	0.42	0.44
8. questions	99.0	87.4	0.46	1.03	92.3	69.2	0.62	0.75
9. one	88.2	82.1	0.68	1.08	69.2	46.2	0.85	0.99
10. them	85.3	*90.5	0.75	1.12	30.8	*80.8	0.50	0.85
11. How	97.1	90.5	1.12	1.17	92.3	92.3	0.27	0.34
12. are	97.1	89.5	1.12	1.07	96.2	92.3	0.00	0.34
13. Well	88.2	81.1	0.84	0.83	96.2	65.4	0.44	0.57
14. answered	98.0	76.8	0.02	0.86	73.1	50.0	0.54	0.77
15. don't	97.1	38.9	1.05	0.18	100.0	80.8	0.00	0.44
16. doctor	100.0	85.3	0.00	0.81	100.0	96.2	0.00	0.42
17. I	99.0	92.6	0.43	0.83	100.0	100.0	0.00	0.00
18. try	94.1	81.1	0.67	0.99	84.6	73.1	0.00	0.48
19. think	93.1	92.6	0.95	0.90	80.8	76.9	0.05	0.50
20. thought	95.1	94.7	0.99	0.65	100.0	84.6	0.00	0.75
21. a	99.0	95.8	0.07	0.62	100.0	100.0	0.00	0.00
22. and	100.0	88.4	0.00	0.64	88.5	88.5	0.91	0.67
23. said	97.1	89.5	0.13	0.95	76.9	*84.6	0.74	0.75
24. I	100.0	94.7	0.00	0.83	100.0	96.2	0.00	0.14
25. now	94.1	82.1	0.65	0.53	76.9	57.7	0.78	0.39

* indicates that the MC-Test item is easier than the corresponding C-Test item (a difference of 5% or greater).

Table B-2

Comparison of Item Statistics for C1S2 and MC1S2

Item	L1 Data				L2 Data			
	Difficulty		Discrimination		Difficulty		Discrimination	
	C1S2	MC1S2	C1S2	MC1S2	C1S2	MC1S2	C1S2	MC1S2
1. the	99.0	98.9	0.99	0.20	92.3	69.2	0.62	0.73
2. is	97.1	98.9	0.98	-0.06	80.8	*100.0	0.88	0.00
3. there	90.2	90.5	0.70	0.56	34.6	30.8	0.11	0.72
4. It	99.0	100.0	0.99	0.00	96.2	92.2	0.57	0.66
5. to	99.0	87.4	0.99	0.24	96.2	53.8	0.04	0.43
6. from	100.0	86.3	0.00	0.47	80.8	46.2	0.70	1.02
7. flashlight	99.0	98.9	0.43	0.16	100.0	92.3	0.00	0.70
8. the	100.0	97.9	0.00	0.78	88.5	69.2	0.48	0.74
9. It	99.0	98.9	0.99	0.63	84.6	*92.3	0.61	0.73
10. the	100.0	98.9	0.00	0.63	88.5	73.1	0.40	0.53
11. room	98.0	97.9	0.81	0.41	65.4	*84.6	0.40	0.40
12. fast	98.0	89.5	0.58	0.75	92.3	53.8	0.22	0.53
13. you	99.0	94.7	0.77	0.66	96.2	88.5	0.57	0.22
14. not	91.2	93.7	0.81	0.70	88.5	76.9	0.14	0.67
15. that	99.0	96.8	0.99	0.56	69.2	*76.9	-0.25	0.48
16. is	99.0	100.0	-0.01	0.00	96.2	96.2	0.57	0.27
17. any	67.6	*81.1	0.76	0.88	50.0	*61.5	0.69	0.30
18. Light	99.0	94.1	0.77	0.94	96.2	92.3	0.87	0.49
19. faster	100.0	95.8	0.00	0.48	96.2	57.7	0.53	0.72
20. anything	100.0	92.6	0.00	0.80	88.5	46.2	0.65	0.71
21. that	82.4	*97.9	0.48	0.33	65.4	*76.9	0.07	0.14
22. has	89.2	93.7	0.57	0.42	42.3	*53.8	0.46	0.67
23. measured	92.2	48.4	0.78	0.40	42.3	15.4	0.29	0.96
24. travels	98.0	92.6	0.78	0.55	100.0	73.1	0.00	0.78
25. fast	100.0	95.8	0.00	0.62	96.2	57.7	0.38	0.67

* indicates that the MC-Test item is easier than the corresponding C-Test item (a difference of 5% or greater).

Table B-3

Comparison of Item Statistics for C1S3 and MC1S3

	L1 Data				L2 Data			
Item	Difficulty		Discrimination		Difficulty		Discrimination	
	C1S3	MC1S3	C1S3	MC1S3	C1S3	MC1S3	C1S3	MC1S3
1. has	100.0	95.8	0.00	0.41	84.6	65.4	-0.01	0.88
2. water	84.3	75.8	0.77	0.54	65.4	26.9	0.13	0.96
3. last	97.1	89.5	0.97	0.33	65.4	•76.9	0.32	0.57
4. several	95.1	98.9	0.27	0.63	92.3	96.2	0.13	0.47
5. At	96.1	95.8	0.91	0.41	73.1	53.8	0.38	0.32
6. one	87.3	62.1	0.32	0.75	53.8	26.9	0.61	0.69
7. however	99.0	68.4	0.16	0.81	88.5	57.7	0.76	0.68
8. body	96.1	83.2	1.24	0.94	88.5	84.6	0.55	0.81
9. only	94.1	75.8	0.79	0.89	46.2	•57.7	0.55	0.89
10. them	99.0	93.7	0.77	1.20	100.0	80.8	0.00	-0.08
11. in	97.1	95.8	0.73	0.65	53.8	•76.9	0.53	0.43
12. lungs	98.0	87.4	0.81	1.03	30.8	•38.5	0.73	0.79
13. last	97.1	83.3	0.97	0.81	53.8	57.7	0.52	0.73
14. three	95.1	87.4	0.60	0.92	76.9	53.8	0.43	0.70
15. four	97.1	87.4	0.24	0.71	88.5	53.8	-0.09	0.50
16. Fortunately	92.2	51.6	0.92	0.65	57.7	23.1	0.21	0.64
17. is	100.0	97.9	0.00	0.78	96.2	88.5	0.57	0.49
18. difficult	54.9	•85.3	0.53	0.90	57.7	57.7	0.28	0.55
19. us	99.0	80.0	0.53	0.88	84.6	69.2	0.40	0.54
20. acquire	85.3	40.0	0.90	0.74	23.1	11.5	0.71	0.86
21. oxygen	100.0	93.7	0.00	1.20	100.0	76.9	0.00	0.09
22. need	99.0	82.1	0.99	0.79	100.0	73.1	0.00	0.46
23. a	94.1	95.8	0.73	0.77	92.3	88.5	0.37	0.18
24. we	100.0	92.6	0.00	0.63	100.0	69.2	0.00	0.83
25. only	94.1	82.1	0.69	0.84	38.5	•57.7	0.50	0.89

• indicates that the MC-Test item is easier than the corresponding C-Test item (a difference of 5% or greater).

Table B-4

Comparison of Item Statistics for C1S4 and MC1S4

Item	L1 Data				L2 Data			
	Difficulty		Discrimination		Difficulty		Discrimination	
	C1S4	MC1S4	C1S4	MC1S4	C1S4	MC1S4	C1S4	MC1S4
1. reason	96.1	75.8	0.92	0.56	69.2	26.9	0.46	0.73
2. this	85.3	97.9	0.76	0.87	34.6	*61.5	0.47	0.60
3. that	91.2	92.6	1.14	0.92	38.5	*65.4	0.71	0.68
4. group	91.2	86.3	0.94	1.18	88.5	84.6	0.34	0.85
5. lived	29.4	*41.1	0.26	0.68	15.4	15.4	0.03	0.73
6. isolation	89.2	33.7	0.82	0.64	46.2	7.7	0.31	0.90
7. reason	97.1	76.8	0.65	0.64	65.4	26.9	0.48	0.73
8. that	97.1	88.4	1.32	0.86	53.8	57.7	0.41	0.76
9. language	97.1	87.4	1.11	0.96	100.0	84.6	0.00	0.32
10. up	83.3	*96.8	0.72	0.17	73.1	*96.2	0.62	0.36
11. under	78.4	29.5	0.61	0.69	57.7	7.7	0.49	0.90
12. conditions	97.1	37.9	0.87	0.57	84.6	11.5	0.59	0.59
13. over	100.0	46.3	0.00	0.69	96.2	23.1	0.04	0.83
14. centuries	97.1	29.5	1.15	0.69	34.6	7.7	0.63	0.90
15. groups	92.2	81.1	1.01	0.99	96.2	76.9	0.08	0.76
16. in	94.1	89.5	1.14	0.67	65.4	*80.8	0.45	0.11
17. with	96.1	93.7	1.28	0.71	88.5	80.8	0.72	0.36
18. groups	92.2	83.2	0.99	1.11	92.3	84.6	0.08	0.85
19. this	64.7	*93.7	0.28	0.90	38.5	*61.5	0.37	0.54
20. their	45.1	*87.4	0.50	0.91	26.9	*73.1	0.23	0.90
21. grew	92.2	89.5	0.80	0.86	50.0	46.2	0.22	0.43
22. and	99.0	81.1	0.99	0.54	80.8	61.5	0.49	0.67
23. alike	71.6	54.7	0.63	0.51	3.8	3.8	0.96	0.92
24. other	97.1	70.5	0.67	0.75	84.6	50.0	0.31	0.51
25. however	98.0	67.4	0.83	0.82	88.5	53.8	0.76	0.65

* indicates that the MC-Test item is easier than the corresponding C-Test item (a difference of 5% or greater).

Table B-5

Comparison of Item Statistics for C2S1 and MC2S1

Item	L1 Data				L2 Data			
	Difficulty		Discrimination		Difficulty		Discrimination	
	C2S1	MC2S1	C2S1	MC2S1	C2S1	MC2S1	C2S1	MC2S1
1. a	100.0	100.0	0.00	0.00	100.0	96.0	0.00	-0.11
2. doctor	100.0	100.0	0.00	0.00	100.0	100.0	0.00	0.00
3. did	98.0	81.1	0.50	0.63	92.3	88.5	0.64	0.03
4. know	99.0	96.7	0.08	0.54	84.6	57.7	0.31	0.23
5. so	100.0	94.4	0.00	0.57	100.0	80.8	0.00	0.35
6. first	65.7	*73.3	0.68	0.54	36.6	*69.2	0.36	0.33
7. some	96.1	85.6	1.12	0.56	76.9	80.8	0.45	0.31
8. and	99.0	70.0	0.45	0.85	69.2	73.1	0.59	0.19
9. of	99.0	95.6	0.25	0.44	73.1	*92.3	0.80	0.08
10. was	93.1	80.0	0.82	0.71	57.7	*76.9	0.42	0.31
11. old	94.1	71.1	0.56	0.67	92.3	65.4	0.81	0.61
12. you	98.0	97.8	0.12	0.42	100.0	96.2	0.00	0.74
13. she	90.2	91.1	0.94	0.55	88.5	88.5	0.39	0.13
14. I	100.0	100.0	0.00	0.00	96.2	100.0	-0.19	0.00
15. remember	100.0	100.0	0.00	0.00	92.3	96.2	0.99	0.06
16. but	100.0	85.6	0.00	0.77	92.3	61.5	0.57	0.70
17. will	99.0	93.3	0.14	0.13	84.6	*100.0	0.11	0.00
18. to	99.0	98.9	0.37	0.44	100.0	100.0	0.00	0.00
19. She	92.2	88.9	1.03	0.94	92.3	88.5	0.52	0.52
20. for	100.0	91.1	0.00	0.75	61.5	*76.9	0.78	0.34
21. minute	100.0	91.1	0.00	0.70	88.5	80.8	1.01	0.73
22. then	95.1	87.8	0.61	0.78	76.9	*84.6	0.76	0.44
23. Yes	93.1	84.4	0.60	0.67	61.5	*73.1	0.76	0.50
24. remember	100.0	98.9	0.00	0.44	96.2	96.2	0.90	0.06
25. doctor	100.0	96.7	0.00	0.51	100.0	100.0	0.00	0.00

* indicates that the MC-Test item is easier than the corresponding C-Test item (a difference of 5% or greater).

Table B-6

Comparison of Item Statistics for C2S2 and MC2S2

Item	L1 Data				L2 Data			
	Difficulty		Discrimination		Difficulty		Discrimination	
	C2S2	MC2S2	C2S2	MC2S2	C2S2	MC2S2	C2S2	MC2S2
1. ligth	100.0	100.0	0.00	00.0	100.0	100.0	0.00	0.00
2. not	88.2	84.4	0.42	0.63	76.9	73.1	0.55	0.03
3. instantly	94.1	80.0	0.71	0.69	92.3	88.5	0.73	0.43
4. has	84.3	80.0	0.43	0.85	65.4	34.6	0.66	0.65
5. travel	97.1	87.8	0.37	0.86	76.9	80.8	0.64	0.28
6. your	100.0	84.4	0.00	0.74	88.5	42.3	0.64	0.49
7. to	100.0	94.4	0.00	0.53	96.2	92.3	0.90	0.17
8. wall	100.0	95.6	0.00	0.58	88.5	*100.0	0.88	0.00
9. crosses	67.6	43.3	0.48	0.84	42.3	23.1	0.76	0.75
10. entire	88.2	80.0	0.94	0.71	30.8	23.1	0.82	0.90
11. so	100.0	86.7	0.00	0.88	96.2	23.1	0.44	0.71
12. that	98.0	96.7	1.11	0.70	50.0	*69.2	0.39	0.93
13. are	97.1	71.1	0.53	0.61	100.0	65.4	0.00	0.57
14. aware	94.1	47.8	1.11	0.86	46.2	19.2	0.82	0.97
15. it	97.1	94.4	0.48	0.75	92.3	88.5	0.99	0.62
16. taking	97.1	35.6	0.84	0.71	76.9	15.4	0.82	0.76
17. time	96.1	67.8	0.98	0.54	80.8	50.0	0.85	0.58
18. travels	97.1	90.0	0.84	0.94	88.5	76.9	0.72	0.65
19. than	79.4	*98.9	0.28	0.64	34.6	*92.3	-0.07	0.55
20. else	98.0	87.8	0.87	0.77	88.5	73.1	0.73	0.68
21. man	88.2	56.7	0.96	0.81	50.0	19.2	0.28	0.74
22. ever	89.2	76.7	0.58	0.90	46.2	38.5	0.67	0.75
23. It	98.0	97.8	0.59	0.49	80.8	84.6	0.53	0.75
24. so	100.0	87.8	0.00	0.76	96.2	34.6	0.44	0.54
25. that	98.0	96.7	1.11	0.69	38.5	*69.2	0.53	0.86

* indicates that the MC-Test item is easier than the corresponding C-Test item (a difference of 5% or greater).

Table B-7

Comparison of Item Statistics for C2S3 and MC2S3

Item	L1 Data				L2 Data			
	Difficulty		Discrimination		Difficulty		Discrimination	
	C2S3	MC2S3	C2S3	MC2S3	C2S3	MC2S3	C2S3	MC2S3
1. enough	99.0	100.0	0.82	0.00	92.3	88.5	0.70	0.20
2. to	100.0	98.9	0.00	0.40	96.2	92.3	0.90	0.38
3. for	100.0	98.9	0.00	0.10	88.5	84.6	0.72	0.29
4. days	100.0	98.9	0.00	0.23	80.8	84.6	0.87	0.76
5. any	65.7	58.9	0.60	0.60	46.2	42.3	0.18	0.67
6. time	96.1	66.7	0.68	0.56	84.6	42.3	0.79	0.62
7. the	98.0	93.3	0.81	0.59	84.6	76.9	0.64	0.75
8. has	94.1	92.2	0.02	0.57	76.9	73.1	0.66	0.90
9. enough	99.0	98.9	0.82	0.40	96.2	•84.6	0.18	0.50
10. stored	89.2	70.0	0.69	0.63	30.8	•38.5	0.80	0.76
11. the	94.1	97.8	0.64	0.55	76.9	84.6	0.60	0.95
12. to	100.0	96.7	0.00	0.61	92.3	76.9	1.02	0.85
13. for	100.0	97.8	0.00	0.36	80.8	80.8	0.90	0.63
14. or	98.0	96.7	0.75	0.63	69.2	•96.2	0.71	0.79
15. minutes	98.0	91.1	0.30	0.73	76.9	57.7	0.55	0.71
16. it	100.0	100.0	0.00	0.00	92.3	96.2	0.99	0.79
17. not	90.2	85.6	0.31	0.56	76.9	80.8	0.49	0.35
18. for	100.0	98.9	0.00	0.40	96.2	88.5	0.90	0.21
19. to	100.0	100.0	0.00	0.00	96.2	76.9	0.90	0.80
20. the	99.0	96.7	0.82	0.67	80.8	76.9	0.80	0.73
21. we	100.0	91.1	0.00	0.33	96.2	73.1	0.90	0.82
22. As	68.6	•75.6	0.57	0.77	42.3	34.6	0.37	0.71
23. rule	72.5	45.6	0.79	0.68	53.8	26.9	0.65	0.64
24. need	100.0	87.8	0.00	0.47	92.3	80.8	0.99	0.84
25. to	98.0	93.3	0.04	0.69	96.2	84.6	0.90	0.98

• indicates that the MC-Test item is easier than the corresponding C-Test item (a difference of 5% or greater).

Table B-8

Comparison of Item Statistics for C2S4 and MC2S4

Item	L1 Data				L2 Data			
	Difficulty		Discrimination		Difficulty		Discrimination	
	C2S4	MC2S4	C2S4	MC2S4	C2S4	MC2S4	C2S4	MC2S4
1. for	100.0	98.9	0.00	0.09	88.5	73.1	1.04	0.33
2. is	99.0	100.0	0.54	0.00	69.2	*84.6	0.49	1.05
3. each	99.0	82.2	0.23	0.17	76.9	69.2	1.15	0.86
4. often	87.3	44.4	0.54	0.53	53.8	19.2	0.32	0.83
5. in	96.1	93.3	0.85	0.56	61.5	65.4	0.38	0.50
6. Another	99.0	62.2	0.73	0.61	96.2	46.2	0.67	0.46
7. is	98.0	98.9	0.32	0.54	61.5	*88.5	~~0.91~~	1.04
8. each	98.0	80.0	0.34	0.29	73.1	69.2	1.09	0.86
9. grew	77.5	*88.9	0.48	0.81	46.2	*76.9	0.27	0.76
10. separately	76.5	38.9	0.81	0.78	38.5	26.9	0.88	0.83
11. different	97.1	67.8	0.58	0.68	61.5	61.5	0.68	0.75
12. Gradually	94.1	27.8	0.70	0.64	46.2	7.7	0.80	0.82
13. many	98.0	93.3	0.63	0.85	80.8	65.4	0.80	0.82
14. some	89.2	80.0	0.65	0.95	80.8	50.0	0.85	0.87
15. came	84.3	44.4	1.01	0.68	53.8	30.8	0.69	0.75
16. contact	79.4	65.6	0.91	0.83	38.5	34.6	0.76	0.88
17. other	95.1	82.2	0.57	0.76	80.8	53.8	0.86	0.84
18. When	94.1	50.0	1.03	0.80	61.5	42.33	0.55	0.54
19. happened	84.3	64.4	0.87	0.79	57.7	26.9	0.85	0.79
20. languages	91.2	73.3	0.61	0.79	84.6	53.8	0.42	0.85
21. more	96.1	66.7	0.27	0.82	92.3	38.5	1.02	0.79
22. more	97.1	65.6	0.39	0.79	92.3	38.5	1.02	0.79
23. In	81.4	77.8	0.94	0.57	53.8	53.8	0.69	0.76
24. cases	74.5	37.8	0.89	0.69	30.8	11.5	0.62	0.54
25. a	94.1	88.9	0.64	0.74	80.8	73.1	0.84	0.91

* indicates that the MC-Test item is easier than the corresponding C-Test item (a difference of 5% or greater).

APPENDIX C

CROSS-TABULATIONS OF ITEM DISCRIMINATION

Table C-1

Cross-tabulation of Item Discrimination for C1S1 and MC1S1 (L1 Data)

		MC1S1			
	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C1S1	>.75	15	3	20	11,12 13,19
	.51 to .75			1,2 25	9,10 18
	.26 to .50				5,8 17
	<.26			7,21 22	4,6,14 16,23,24

Table C-2

Cross-tabulation of Item Discrimination for C1S2 and MC1S2 (L1 Data)

		MC1S2			
	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C1S2	>.75	1,2,4 5	11,23	9,13,14 15,24	17,18
	.51 to .75		22	3,12	
	.26 to .50	7	21		
	<.26	16	6,19	10,25	8,20

Note: Numerals in the cells represent test item numbers. This applies to Tables C-1 through C-16.

Table C-3

Cross-tabulation of Item Discrimination for C1S3 and MC1S3 (L1 Data)

MC1S3

	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C1S3	>.75		3,5	2,16,20	8,9,10 12,13,22
	.51 to .75			11	14,18,19 23,25
	.26 to .50			4,6	
	<.26		1	15,24	7,17,21

Table C-4

Cross-tabulation of Item Discrimination for C1S4 and MC1S4 (L1 Data)

MC1S4

	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C1S4	>.75			1,6,12 14,16,17 22	2,3,4 8,9,15 18,21,25
	.51 to .75	10		7,11,23 24	
	.26 to .50			5	19,20
	<.26			13	

Table C-5

Cross-tabulation of Item Discrimination for C2S1 and MC2S1 (L1 Data)

MC2S1

	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C2S1	>.75			7,10,13	19
	.51 to .75			6,11,23	22
	.26 to .50		18	3	8
	<.26	1,2,14 15,17	9,12,24	4,5,20 21,25	16

Table C-6

Cross-tabulation of Item Discrimination for C2S2 and MC2S2 (L1 Data)

MC2S2

	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C2S2	>.75			10,12,16 17,25	14,18,20 21
	.51 to .75		23	3,13	22
	.26 to .50			2,15,19	4,5,9
	<.26	1		6,7,8	11,24

Table C-7

Cross-tabulation of Item Discrimination for C2S3 and MC2S3 (L1 Data)

		MC2S3			
	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
	>.75	1	9	7,20,23	
C2S3				5,6,10 11,14	22
	.26 to .50			15,17	
	<.26	3,4,16 19	2,13,18 21,24	8,12,25	

Table C-8

Cross-tabulation of Item Discrimination for C2S4 and MC2S4 (L1 Data)

		MC2S4			
	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
	>.75			5,15,23 24	10,16,18 19
C2S4	.51 to .75	2		4,6,11 12,25	13,14,17 20
	.26 to .50		8	7	9,21,22
	<.26	1,3			

Table C-9

Cross-tabulation of Item Discrimination for C1S1 and MC1S1 (L2 Data)

		MC1S1			
	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C1S1	>.75		25	2,22	9
	.51 to .75	1	3	8,23	14
	.26 to .50		7,11	13	4,10
	<.26	17,21 24	6,12,15 16,18,19	5,20	

Table C-10

Cross-tabulation of Item Discrimination for C1S2 and MC1S2 (L2 Data)

		MC1S2			
	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C1S2	>.75	2	18		
	.51 to .75	13	16,17	1,4,9 19,20	6
	.26 to .50		11	8,10,22 25	23
	<.26	21	5,15	3,7,12 14	24

Table C-11

Cross-tabulation of Item Discrimination for C1S3 and MC1S3 (L2 Data)

MC1S3

	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C1S3	>.75			7	
	.51 to .75		11,17	6,13	8,9,12 20
	.26 to .50	23	5	3,14,18 19	25
	<.26	10,21	4,15,22	16	1,2,24

Table C-12

Cross-tabulation of Item Discrimination for C1S4 and MC1S4 (L2 Data)

MC1S4

	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C1S4	>.75			25	23
	.51 to .75		10,17	3,12	14
	.26 to .50	16		1,2,7 19,22,24	4,6,8 11
	<.26		9,21	5	13,15,18 20

Table C-13

Cross-tabulation of Item Discrimination for C2S1 and MC2S1 (L2 Data)

		MC2S1			
	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C2S1	>.75	9,15,24	20,22,23	11,21	
	.51 to .75	3,8		16,19	
	.26 to .50	4,13	6,7,10		
	<.26	1,2,14 17,18,25	5	12	

Table C-14

Cross-tabulation of Item Discrimination for C2S2 and MC2S2 (L2 Data)

		MC2S2			
	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C2S2	>.75	7,8		9,15,17	10,14,16
	.51 to .75	2	3,5,6	4,18,20 22,23	25
	.26 to .50			11,21,24	12
	<.26	1		13,19	

Table C-15

Cross-tabulation of Item Discrimination for C2S3 and MC2S3 (L2 Data)

		MC2S3			
	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C2S3	>.75	18	2	6,13,20	4,10,12 16,19,21 24,25
	.51 to .75	1	3	7,15,23	8,11,14
	.26 to .50		17	22	
	<.26		9	5	

Table C-16

Cross-tabulation of Item Discrimination for C2S4 and MC2S4 (L2 Data)

		MC2S4			
	Discrimination Interval	<.26	.26 to .50	.51 to .75	>.75
C2S4	>.75		1		3,7,8 10,12,13 14,16,17 19,21,22,25
	.51 to .75		6	11,15,18 24	23
	.26 to .50		5		2,4,9 20
	<.26				

APPENDIX D

FACTOR LOADING MATRICES

Table D-1: Varimax Rotated Factor Matrix of C1S1

	H**2	1	2	3	4	5
TEST 1	0.736	0.101	0.137	0.101	0.833	0.053
TEST 2	0.579	0.122	0.046	0.646	-0.044	0.377
TEST 3	0.532	0.635	-0.100	-0.341	0.016	0.038
TEST 4	0.008	-0.036	-0.003	0.026	-0.037	-0.065
TEST 5	0.400	-0.040	-0.019	0.629	-0.044	-0.036
TEST 6	0.856	-0.127	0.302	-0.158	0.832	0.177
TEST 7	0.730	0.545	-0.114	0.577	0.245	-0.165
TEST 8	0.701	0.427	-0.059	0.515	0.498	0.053
TEST 9	0.841	0.874	0.170	0.173	0.121	0.061
TEST 10	0.841	0.874	0.170	0.173	0.121	0.061
TEST 11	0.726	0.291	0.114	0.115	0.174	0.765
TEST 12	0.616	-0.111	-0.119	0.109	0.002	0.760
TEST 13	0.764	0.796	0.217	0.183	-0.095	0.204
TEST 14	0.577	0.514	0.432	0.279	-0.179	0.129
TEST 15	0.857	0.071	0.889	-0.011	0.246	0.027
TEST 16	0.862	0.345	0.815	-0.120	0.247	0.055
TEST 17	0.839	0.446	0.736	-0.080	0.287	0.102
TEST 18	0.259	0.025	-0.172	0.264	0.244	-0.316
TEST 19	0.390	-0.202	0.506	0.209	-0.195	-0.110
TEST 20	0.550	0.343	0.364	0.544	0.023	-0.056
% COM VAR.		31.034	22.676	17.294	16.597	12.399
% TOT VAR.		19.652	14.359	10.951	10.510	7.852

Table D-2: Varimax Rotated Factor Matrix of MC1S1

	H**2	1	2	3	4	5
TEST 1	0.705	-0.040	0.697	0.040	0.075	0.458
TEST 2	0.548	0.032	0.021	0.623	0.394	0.057
TEST 3	0.694	-0.077	0.095	0.143	0.800	0.132
TEST 4	0.701	0.180	0.701	0.183	0.356	0.130
TEST 5	0.641	0.174	0.685	0.330	0.183	-0.008
TEST 6	0.633	-0.051	0.769	0.167	-0.044	-0.094
TEST 7	0.678	0.049	0.222	0.780	0.121	-0.058
TEST 8	0.867	0.132	0.456	0.793	0.111	0.016
TEST 9	0.748	0.356	0.487	0.607	0.090	-0.083
TEST 10	0.703	0.337	0.496	0.517	-0.031	0.274
TEST 11	0.824	0.446	0.724	0.306	-0.078	-0.033
TEST 12	0.759	0.671	0.494	0.202	-0.121	-0.097
TEST 13	0.502	0.477	0.058	0.448	-0.138	0.226
TEST 14	0.591	0.380	0.335	0.525	-0.135	0.201
TEST 15	0.678	0.084	0.063	0.041	0.098	0.810
TEST 16	0.369	0.546	0.261	0.048	0.009	-0.007
TEST 17	0.662	0.775	0.219	0.067	0.020	-0.095
TEST 18	0.589	0.580	0.347	0.357	0.065	0.023
TEST 19	0.826	0.784	0.333	0.001	0.250	-0.196
TEST 20	0.707	0.755	-0.095	0.125	0.092	0.321
TEST 21	0.642	0.777	-0.112	0.011	0.116	0.112
TEST 22	0.761	0.592	0.143	0.120	0.608	-0.079
TEST 23	0.725	0.699	0.082	0.474	-0.075	0.011
TEST 24	0.801	0.777	-0.127	0.357	-0.127	0.194
TEST 25	0.370	0.595	0.104	0.065	-0.020	0.037
% COM VAR.		36.955	24.515	21.449	9.258	7.824
% TOT VAR.		24.718	16.398	14.347	6.192	5.233

Table D-3: Varimax Rotated Factor Matrix of C1S2

		H**2	1	2	3	4
TEST	1	0.997	0.995	-0.014	0.086	-0.000
TEST	2	0.597	-0.016	0.621	-0.027	0.459
TEST	3	0.712	-0.028	0.376	-0.070	0.752
TEST	4	0.997	0.995	-0.014	0.086	-0.000
TEST	5	0.997	0.995	-0.014	0.086	-0.000
TEST	6	0.044	-0.010	-0.042	-0.050	0.199
TEST	7	0.997	0.995	-0.014	0.086	-0.000
TEST	8	0.630	-0.003	0.778	-0.024	0.155
TEST	9	0.609	-0.001	0.773	0.105	-0.030
TEST	10	0.933	-0.005	0.959	0.104	-0.039
TEST	11	0.487	0.306	0.341	0.503	0.155
TEST	12	0.997	0.995	-0.014	0.086	-0.000
TEST	13	0.038	0.008	0.026	-0.104	-0.163
TEST	14	0.637	0.115	0.135	0.414	0.659
TEST	15	0.933	-0.005	0.959	0.104	-0.039
TEST	16	0.609	0.156	-0.163	0.683	0.303
TEST	17	0.688	-0.121	0.238	0.756	-0.212
TEST	18	0.553	0.305	0.266	0.623	0.003
TEST	19	0.657	-0.043	0.724	0.355	-0.064
		% COM VAR.	39.580	33.457	15.765	11.198
		% TOT VAR.	27.311	23.086	10.878	7.727

Table D-4: Varimax Rotated Factor Matrix of MC1S2

		H**2	1	2	3	4	5
TEST	1	0.021	0.002	0.017	-0.046	-0.114	-0.076
TEST	2	0.910	-0.047	-0.020	-0.040	0.023	0.952
TEST	3	0.671	0.524	-0.057	0.388	-0.087	0.485
TEST	4	0.737	0.186	-0.017	-0.101	0.828	-0.083
TEST	5	0.790	0.133	0.305	0.196	0.800	-0.038
TEST	6	0.835	-0.118	0.044	0.895	0.128	0.040
TEST	7	0.664	0.097	0.754	-0.090	0.274	-0.052
TEST	8	0.924	0.009	0.960	-0.033	-0.009	0.048
TEST	9	0.924	0.009	0.960	-0.033	-0.009	0.048
TEST	10	0.954	-0.027	0.668	-0.052	0.010	0.710
TEST	11	0.552	0.544	0.338	0.368	-0.025	0.071
TEST	12	0.654	0.545	0.503	0.010	-0.323	-0.004
TEST	13	0.679	0.397	0.523	0.494	-0.060	0.000
TEST	14	0.798	-0.088	0.665	0.589	0.027	0.010
TEST	15	0.447	0.587	0.282	0.070	-0.060	-0.119
TEST	16	0.781	0.762	0.438	-0.027	0.070	0.054
TEST	17	0.445	0.590	-0.059	-0.095	0.273	0.102
TEST	18	0.708	0.810	-0.025	0.046	0.213	-0.055
TEST	19	0.697	0.089	-0.059	0.824	0.088	-0.013
TEST	20	0.529	0.280	-0.070	0.668	-0.009	-0.007
TEST	21	0.218	0.336	-0.131	0.214	-0.097	-0.180
TEST	22	0.693	0.785	-0.054	0.042	0.267	-0.013
TEST	23	0.829	0.892	-0.077	0.155	-0.045	0.041
		% COM VAR.	30.106	28.110	19.117	11.418	11.249
		% TOT VAR.	20.238	18.896	12.851	7.676	7.562

Table D-5: Varimax Rotated Factor Matrix of C1S3

	H**2	1	2	3	4	5
TEST 1	0.550	0.103	0.296	0.379	0.488	0.264
TEST 2	0.915	0.494	-0.225	0.318	0.719	-0.039
TEST 3	0.571	-0.036	0.225	-0.280	0.662	-0.047
TEST 4	0.549	0.406	0.110	0.595	-0.134	-0.002
TEST 5	0.554	-0.065	-0.036	0.076	-0.081	0.732
TEST 6	0.305	-0.006	0.155	-0.121	0.031	0.515
TEST 7	0.816	0.303	0.239	0.758	0.304	0.005
TEST 8	0.912	0.074	0.896	0.134	0.078	0.283
TEST 9	0.852	0.869	-0.139	0.259	0.102	-0.019
TEST 10	0.828	0.768	0.463	-0.079	0.132	-0.033
TEST 11	0.869	0.905	0.222	0.011	-0.025	0.029
TEST 12	0.915	0.494	-0.225	0.318	0.719	-0.039
TEST 13	0.273	0.503	-0.073	0.122	-0.018	0.004
TEST 14	0.043	0.080	-0.030	-0.021	-0.177	-0.061
TEST 15	0.588	0.211	-0.136	0.684	0.112	0.211
TEST 16	0.435	0.134	0.206	-0.011	0.343	0.506
TEST 17	0.420	-0.052	-0.201	0.180	-0.015	0.587
TEST 18	0.537	0.047	0.386	0.051	0.163	0.597
TEST 19	0.713	-0.153	0.421	0.706	-0.114	-0.033
TEST 20	0.660	-0.134	0.489	0.517	0.353	-0.109
TEST 21	0.814	0.071	0.880	0.177	0.047	0.034
	% COM VAR.	25.167	21.856	21.187	16.712	15.078
	% TOT VAR.	15.722	13.653	13.235	10.440	9.419

Table D-6: Varimax Rotated Factor Matrix of MC1S3

	H**2	1	2	3	4
TEST 1	0.315	-0.080	0.306	-0.146	0.439
TEST 2	0.457	0.161	0.531	-0.320	0.216
TEST 3	0.789	-0.015	0.101	0.033	0.882
TEST 4	0.732	0.141	0.060	0.842	-0.016
TEST 5	0.530	-0.076	0.135	0.709	0.065
TEST 6	0.499	0.120	0.647	0.169	0.192
TEST 7	0.309	0.324	0.431	0.118	-0.063
TEST 8	0.453	0.395	0.490	0.148	0.186
TEST 9	0.588	0.233	0.678	0.130	0.238
TEST 10	0.883	0.815	0.250	0.293	0.265
TEST 11	0.539	0.652	0.149	-0.213	-0.215
TEST 12	0.612	0.659	0.290	0.182	0.246
TEST 13	0.787	0.522	0.097	0.256	0.663
TEST 14	0.577	0.638	0.278	0.297	0.068
TEST 15	0.528	0.590	0.259	0.267	-0.205
TEST 16	0.494	0.102	0.686	0.056	-0.097
TEST 17	0.701	0.794	0.130	-0.212	-0.092
TEST 18	0.496	0.547	0.372	0.215	0.111
TEST 19	0.471	0.412	0.485	0.172	0.191
TEST 20	0.591	0.004	0.762	0.084	0.049
TEST 21	0.883	0.815	0.250	0.293	0.265
TEST 22	0.359	0.351	0.413	0.253	-0.039
TEST 23	0.664	0.800	0.058	-0.137	-0.043
TEST 24	0.548	0.716	0.052	-0.140	0.110
TEST 25	0.567	0.287	0.659	-0.165	0.149
	% COM VAR.	42.244	28.898	14.978	13.879
	% TOT VAR.	24.281	16.610	8.609	7.978

Table D-7: Varimax Rotated Factor Matrix of C1S4

	H**2	1	2	3	4	5
TEST 1	0.928	0.115	-0.125	0.886	0.337	0.023
TEST 2	0.578	-0.108	0.470	0.224	0.266	0.474
TEST 3	0.681	0.310	0.231	0.713	0.123	0.087
TEST 4	0.556	0.628	0.167	-0.090	0.351	0.053
TEST 5	0.245	0.013	-0.110	0.079	0.149	0.452
TEST 6	0.611	0.397	0.333	0.209	0.066	0.542
TEST 7	0.849	-0.020	-0.067	0.866	0.132	0.278
TEST 8	0.800	0.172	0.449	0.484	0.458	-0.353
TEST 9	0.854	0.742	0.496	-0.075	0.001	0.226
TEST 10	0.488	0.046	0.112	0.228	0.649	0.013
TEST 11	0.604	0.107	0.097	-0.032	0.756	0.100
TEST 12	0.595	0.513	0.574	-0.044	-0.007	-0.024
TEST 13	0.825	0.500	0.261	0.533	0.285	-0.375
TEST 14	0.779	0.833	0.080	0.228	0.163	-0.017
TEST 15	0.639	0.572	0.286	0.243	0.391	-0.131
TEST 16	0.751	0.535	0.441	0.012	0.484	-0.186
TEST 17	0.788	0.844	0.079	0.216	0.140	-0.045
TEST 18	0.230	-0.062	0.179	0.379	-0.205	0.092
TEST 19	0.323	0.150	0.005	0.061	0.463	0.287
TEST 20	0.494	0.091	0.634	0.080	0.202	0.192
TEST 21	0.804	0.269	0.796	0.081	0.112	-0.280
TEST 22	0.315	0.250	0.044	0.204	0.411	0.200
TEST 23	0.851	0.776	-0.096	0.211	-0.099	0.429
TEST 24	0.736	0.082	0.851	0.060	-0.003	-0.040
% COM VAR.		29.899	22.223	20.481	16.795	10.602
% TOT VAR.		19.088	14.188	13.075	10.722	6.769

Table D-8: Varimax Rotated Factor Matrix of MC1S4

	H**2	1	2	3	4	5
TEST 1	0.841	0.158	0.092	0.184	0.875	0.084
TEST 2	0.668	-0.274	0.017	0.711	0.286	-0.071
TEST 3	0.717	0.076	-0.042	0.779	0.279	0.156
TEST 4	0.813	0.283	0.234	0.796	-0.082	0.194
TEST 5	0.602	0.109	0.728	0.138	0.195	0.053
TEST 6	0.618	0.072	0.741	0.069	0.206	0.129
TEST 7	0.845	0.176	0.167	0.183	0.866	0.053
TEST 8	0.725	0.352	0.151	0.512	0.545	-0.135
TEST 9	0.631	0.326	0.269	0.563	0.228	-0.290
TEST 10	0.840	0.507	0.138	-0.107	0.423	-0.611
TEST 11	0.741	0.041	0.853	0.078	0.078	0.005
TEST 12	0.640	0.103	0.754	0.175	-0.094	-0.147
TEST 13	0.689	0.132	0.782	0.240	-0.026	-0.048
TEST 14	0.673	0.060	0.794	0.077	0.101	0.148
TEST 15	0.723	0.475	0.275	0.649	0.003	-0.013
TEST 16	0.644	0.774	0.055	0.150	0.121	0.074
TEST 17	0.716	0.824	0.021	0.112	0.136	-0.079
TEST 18	0.827	0.505	0.252	0.712	-0.012	-0.027
TEST 19	0.881	0.915	0.098	0.173	0.070	-0.004
TEST 20	0.698	0.678	0.200	0.379	0.175	0.155
TEST 21	0.736	0.790	0.175	0.163	0.154	0.179
TEST 22	0.589	0.395	0.137	0.158	0.215	0.586
TEST 23	0.599	0.305	0.401	-0.024	0.360	0.464
TEST 24	0.424	0.277	0.270	0.492	0.110	0.139
TEST 25	0.406	0.279	0.371	0.432	0.050	0.041
% COM VAR.		27.983	25.665	23.991	15.039	7.323
% TOT VAR.		19.345	17.743	16.586	10.397	5.063

Table D-9: Varimax Rotated Factor Matrix of C2S1

	H**2	1	2	3	4
TEST 1	0.319	0.218	0.460	-0.121	-0.212
TEST 2	0.403	0.537	-0.236	-0.125	-0.207
TEST 3	0.326	0.113	0.478	0.246	0.156
TEST 4	0.804	0.256	0.855	0.080	-0.032
TEST 5	0.342	0.373	-0.050	-0.062	0.443
TEST 6	0.275	-0.090	-0.024	-0.031	0.515
TEST 7	0.866	0.361	0.104	0.098	0.846
TEST 8	0.661	-0.008	0.104	0.806	0.023
TEST 9	0.395	-0.104	-0.105	0.609	-0.046
TEST 10	0.843	0.869	0.255	0.066	0.135
TEST 11	0.275	-0.090	-0.024	-0.031	0.515
TEST 12	0.554	-0.271	0.682	-0.119	0.041
TEST 13	0.921	0.897	0.286	0.093	0.158
TEST 14	0.698	0.201	0.090	0.805	-0.038
TEST 15	0.333	-0.380	0.334	0.198	-0.194
% COM VAR.		31.564	24.892	23.043	20.501
% TOT VAR.		16.864	13.299	12.312	10.953

Table D-10: Varimax Rotated Factor Matrix of MC2S1

	H**2	1	2	3	4	5	6
TEST 1	0.635	0.075	0.045	0.130	0.721	0.231	-0.192
TEST 2	0.861	-0.029	-0.078	0.080	0.014	0.920	-0.040
TEST 3	0.820	-0.076	0.072	-0.003	0.079	0.893	-0.078
TEST 4	0.647	0.190	0.173	0.429	0.238	0.191	-0.551
TEST 5	0.563	0.327	-0.089	0.588	0.084	0.310	0.007
TEST 6	0.653	0.034	0.229	0.598	0.212	0.365	0.252
TEST 7	0.761	-0.010	-0.165	0.835	-0.108	-0.114	-0.107
TEST 8	0.685	0.079	0.324	0.719	0.196	0.115	-0.072
TEST 9	0.495	0.098	0.247	0.505	0.099	-0.090	0.388
TEST 10	0.713	0.756	0.018	0.356	0.098	-0.056	-0.038
TEST 11	0.758	0.200	0.082	0.021	0.834	-0.122	-0.012
TEST 12	0.597	0.142	0.419	0.127	0.250	0.381	0.422
TEST 13	0.483	0.523	0.359	0.027	-0.204	-0.072	-0.182
TEST 14	0.929	0.926	0.162	0.012	0.198	0.028	0.079
TEST 15	0.801	0.100	0.388	0.112	0.752	-0.030	0.247
TEST 16	0.844	0.170	0.884	0.081	0.160	0.029	-0.003
TEST 17	0.861	0.161	0.877	0.070	0.248	0.017	0.001
TEST 18	0.685	0.037	0.521	0.056	0.597	0.154	0.172
TEST 19	0.575	0.141	0.299	0.326	-0.091	0.491	0.332
TEST 20	0.929	0.926	0.162	0.012	0.198	0.028	0.079
TEST 21	0.723	0.614	-0.015	0.086	0.115	0.023	0.570
% COM VAR.		21.667	18.127	17.743	17.406	16.205	8.852
% TOT VAR.		15.497	12.965	12.690	12.449	11.590	6.331

Table D-11: Varimax Rotated Factor Matrix of C2S2

		H**2	1	2	3	4	5	6
TEST	1	0.756	0.007	-0.087	0.116	0.077	0.107	0.847
TEST	2	0.434	0.310	0.089	0.149	0.545	-0.064	0.073
TEST	3	0.842	0.123	-0.019	-0.131	-0.174	0.882	-0.035
TEST	4	0.713	-0.088	-0.069	-0.035	0.836	0.006	-0.006
TEST	5	0.726	-0.019	0.140	0.277	0.229	0.749	0.124
TEST	6	0.582	0.343	0.632	-0.115	0.194	-0.108	-0.049
TEST	7	0.936	0.931	0.185	0.143	0.047	0.047	-0.103
TEST	8	0.664	-0.210	0.765	0.004	-0.071	0.118	0.125
TEST	9	0.727	0.464	0.701	0.022	-0.054	0.084	0.102
TEST	10	0.645	0.268	0.004	0.627	0.416	0.008	0.089
TEST	11	0.747	0.232	0.630	0.522	0.031	-0.044	-0.145
TEST	12	0.763	0.793	0.049	0.142	0.083	0.066	0.317
TEST	13	0.757	0.488	0.058	-0.562	0.418	0.049	-0.151
TEST	14	0.468	0.003	0.248	-0.339	-0.071	-0.061	0.532
TEST	15	0.849	0.312	0.656	0.555	0.055	-0.026	-0.101
TEST	16	0.659	0.319	0.516	-0.127	0.316	0.401	0.118
TEST	17	0.668	-0.023	0.515	0.140	0.579	0.213	-0.045
TEST	18	0.587	0.458	0.082	0.571	0.015	0.164	-0.131
TEST	19	0.936	0.931	0.185	0.143	0.047	0.047	-0.103
		% COM VAR.	26.995	22.416	14.647	14.368	12.161	9.412
		% TOT VAR.	19.119	15.876	10.374	10.176	8.613	6.666

Table D-12: Varimax Rotated Factor Matrix of MC2S2

		H**2	1	2	3	4	5
TEST	1	0.334	0.546	0.079	-0.029	0.114	0.123
TEST	2	0.524	0.198	-0.225	0.165	0.634	0.068
TEST	3	0.388	0.430	0.069	0.234	0.307	0.220
TEST	4	0.577	0.326	0.237	0.614	0.064	0.181
TEST	5	0.619	0.384	-0.085	0.004	0.343	0.588
TEST	6	0.638	0.214	0.074	0.118	-0.106	0.750
TEST	7	0.626	-0.004	-0.010	0.125	0.308	0.718
TEST	8	0.535	0.664	0.007	0.114	0.281	0.055
TEST	9	0.654	0.282	0.327	-0.168	0.585	0.312
TEST	10	0.838	0.167	0.298	0.154	0.789	0.274
TEST	11	0.820	0.088	0.855	0.082	0.158	0.223
TEST	12	0.451	0.618	0.229	0.070	0.101	0.033
TEST	13	0.646	0.765	0.011	0.144	0.168	0.106
TEST	14	0.505	0.117	0.456	0.499	-0.020	0.185
TEST	15	0.586	0.755	0.018	0.089	0.089	0.004
TEST	16	0.466	0.627	0.257	0.040	-0.071	0.009
TEST	17	0.714	0.190	0.246	0.738	0.105	0.248
TEST	18	0.684	0.112	0.809	0.120	0.023	-0.053
TEST	19	0.480	0.133	0.263	0.466	0.376	0.185
TEST	20	0.501	0.544	-0.055	0.340	0.183	0.229
TEST	21	0.658	0.622	0.014	0.452	0.132	0.220
TEST	22	0.817	-0.005	-0.057	0.828	0.184	-0.308
TEST	23	0.720	0.218	0.296	0.246	0.705	-0.166
TEST	24	0.753	0.139	0.711	0.278	0.228	-0.314
		% COM VAR.	28.479	19.043	19.028	18.195	15.255
		% TOT VAR.	17.244	11.531	11.522	11.018	9.237

Table D-13: Varimax Rotated Factor Matrix of C2S3

		H**2	1	2	3	4	5
TEST	1	0.991	[illegible]	0.076	-0.045	-0.035	-0.008
TEST	2	0.532	0.108	0.694	-0.177	0.021	0.085
TEST	3	0.572	0.527	0.048	0.357	-0.405	0.011
TEST	4	0.787	0.737	0.040	0.481	0.094	-0.040
TEST	5	0.634	-0.026	0.103	-0.068	0.134	0.775
TEST	6	0.991	0.991	0.076	-0.045	-0.035	-0.008
TEST	7	0.524	0.270	0.306	0.252	-0.486	0.239
TEST	8	0.590	-0.004	0.131	0.524	0.539	-0.083
TEST	9	0.853	0.016	0.008	0.917	-0.110	-0.020
TEST	10	0.457	-0.098	0.574	-0.117	-0.262	-0.191
TEST	11	0.486	0.012	0.018	-0.018	0.693	0.070
TEST	12	0.991	0.991	0.076	-0.045	-0.035	-0.008
TEST	13	0.635	0.104	0.726	0.235	-0.040	0.200
TEST	14	0.772	0.115	0.778	0.292	0.254	-0.062
TEST	15	0.612	-0.024	-0.045	0.011	-0.134	0.769
		% COM VAR.	37.273	20.038	16.626	13.112	12.951
		% TOT VAR.	25.906	13.927	11.556	9.113	9.001

Table D-14: Varimax Rotated Factor Matrix of MC2S3

		H**2	1	2	3	4	5	6
TEST	1	0.963	0.980	-0.009	0.025	0.010	-0.027	0.017
TEST	2	0.888	0.004	0.185	-0.078	-0.018	-0.089	0.916
TEST	3	0.486	0.007	0.502	-0.189	-0.056	-0.313	-0.311
TEST	4	0.706	-0.126	0.805	0.027	0.068	0.155	0.111
TEST	5	0.729	-0.106	0.799	0.067	-0.091	0.218	0.139
TEST	6	0.596	-0.074	0.176	0.331	0.668	-0.011	0.055
TEST	7	0.567	-0.023	0.080	0.736	-0.032	0.126	-0.033
TEST	8	0.963	0.980	-0.009	0.025	0.010	-0.027	0.017
TEST	9	0.570	0.208	0.448	0.442	0.012	0.261	-0.248
TEST	10	0.703	0.024	0.050	-0.291	0.625	0.473	-0.003
TEST	11	0.652	0.645	0.136	-0.159	0.074	0.432	-0.002
TEST	12	0.954	0.700	0.125	-0.038	-0.005	-0.083	0.663
TEST	13	0.470	0.654	0.051	0.158	-0.108	-0.055	-0.014
TEST	14	0.620	0.323	0.257	0.642	0.134	0.139	-0.018
TEST	15	0.620	0.284	0.459	-0.228	0.412	-0.202	-0.258
TEST	16	0.963	0.980	-0.009	0.025	0.010	-0.027	0.017
TEST	17	0.710	-0.038	0.023	0.156	0.826	0.008	-0.032
TEST	18	0.411	-0.063	-0.049	0.556	0.241	-0.192	0.018
TEST	19	0.488	0.244	0.577	0.226	0.124	0.077	0.154
TEST	20	0.353	0.099	0.526	0.133	0.181	0.114	0.051
TEST	21	0.609	-0.056	0.147	0.128	-0.139	0.740	-0.041
TEST	22	0.553	-0.004	0.173	0.022	0.196	0.693	-0.058
		% COM VAR.	31.270	19.154	13.149	13.083	12.541	10.804
		% TOT VAR.	20.709	12.685	8.708	8.665	8.305	7.155

Table D-15: Varimax Rotated Factor Matrix of C2S4

		H**2	1	2	3	4
TEST	1	0.866	0.015	-0.001	0.029	0.930
TEST	2	0.812	0.048	-0.017	0.896	-0.087
TEST	3	0.505	-0.137	0.362	0.502	0.321
TEST	4	0.642	-0.096	0.385	0.545	0.433
TEST	5	0.357	-0.171	0.570	0.043	-0.022
TEST	6	0.668	-0.027	-0.084	-0.028	0.812
TEST	7	0.736	0.044	-0.025	0.856	-0.028
TEST	8	0.355	0.039	0.578	-0.081	-0.115
TEST	9	0.425	-0.052	0.565	-0.266	0.180
TEST	10	0.717	0.012	0.091	0.615	0.575
TEST	11	0.402	0.267	0.160	0.173	0.524
TEST	12	0.370	0.510	0.299	-0.118	-0.078
TEST	13	0.347	-0.107	0.548	-0.159	-0.102
TEST	14	0.587	0.257	0.626	0.095	0.346
TEST	15	0.508	0.334	0.464	0.336	0.261
TEST	16	0.681	0.716	0.125	0.385	-0.064
TEST	17	0.631	0.642	0.219	-0.044	0.411
TEST	18	0.606	0.394	0.606	0.285	-0.048
TEST	19	0.663	0.613	0.018	0.440	0.306
TEST	20	0.674	0.811	-0.106	-0.045	-0.052
TEST	21	0.771	0.874	-0.059	-0.038	-0.037
TEST	22	0.512	0.370	0.567	0.115	0.200
TEST	23	0.518	0.299	0.613	0.124	0.193
TEST	24	0.564	0.611	0.127	-0.086	0.409
		% COM VAR.	29.221	24.829	22.983	22.967
		% TOT VAR.	16.944	14.397	13.327	13.318

Table D-16: Varimax Rotated Factor Matrix of MC2S4

		H**2	1	2	3	4	5	6
TEST	1	0.760	-0.066	0.004	-0.046	-0.061	0.028	0.865
TEST	2	0.908	-0.081	0.083	-0.062	0.944	0.016	-0.010
TEST	3	0.688	0.075	0.315	0.110	-0.029	0.754	0.036
TEST	4	0.690	0.026	-0.094	0.795	0.049	0.190	0.102
TEST	5	0.574	-0.009	0.699	0.128	-0.037	0.208	-0.154
TEST	6	0.537	0.077	0.053	0.718	0.086	-0.069	-0.024
TEST	7	0.916	-0.027	0.169	0.161	0.927	-0.003	-0.039
TEST	8	0.510	0.246	0.314	0.561	-0.065	0.155	-0.086
TEST	9	0.650	0.274	0.597	0.045	0.290	0.342	0.125
TEST	10	0.616	0.232	0.688	0.056	0.254	-0.068	-0.130
TEST	11	0.545	0.345	0.476	-0.087	0.224	0.347	0.145
TEST	12	0.711	0.413	0.348	0.565	-0.087	-0.301	-0.051
TEST	13	0.500	0.322	0.260	0.426	0.038	0.152	0.350
TEST	14	0.553	0.076	0.655	0.175	0.055	0.137	0.256
TEST	15	0.682	0.414	0.700	0.106	-0.096	0.008	0.027
TEST	16	0.593	0.501	0.392	0.377	0.012	-0.213	0.008
TEST	17	0.573	0.434	0.550	-0.009	0.237	0.074	0.140
TEST	18	0.666	0.728	0.287	0.063	0.090	0.115	0.167
TEST	19	0.598	0.742	0.199	0.043	-0.033	-0.055	-0.053
TEST	20	0.919	0.818	0.128	0.183	-0.121	0.418	-0.099
TEST	21	0.909	0.826	0.095	0.173	-0.065	0.414	-0.112
TEST	22	0.631	0.595	0.013	0.213	0.022	-0.032	0.480
TEST	23	0.500	0.395	0.345	0.054	0.176	0.342	0.273
TEST	24	0.729	0.674	0.262	0.299	-0.098	-0.326	0.026
		% COM VAR.	28.946	23.109	15.211	13.259	10.595	8.880
		% TOT VAR.	19.246	15.365	10.114	8.816	7.045	5.905

APPENDIX E

C-TEST ERROR CODING SHEET

C-TEST ERROR CODING SHEET

Name: Thavone Form: [✓ C1] MC1 C2 MC2

Subtest	Text	Error	Error Type					
			Substitution				No Response	Unintelligible
			Grammatical Function		Meaning Change			
			Same	Different	Little or No	High		
S 1	her	how		✓		✓		
	them	this		✓	✓			
S 2	It	Is		✓		✓		
	not	now		✓		✓		
	that	this		✓		✓		
	any	all	✓			✓		
	that	this		✓		✓		
	has	how		✓		✓		
S 3	has	had	✓		✓			
	last	laid	✓			✓		
	several	seventh		✓		✓		
	At	An		✓	✓			
	in	is		✓		✓		
	fungs	futin						✓
	last	laid	✓			✓		
	us	–					✓	
	acquire	acqupid						✓
S 4	isolation	–					✓	
	that	this		✓		✓		
	up	un		✓		✓		
	centuries	centences	✓			✓		
	groups	growed		✓		✓		
	in	is		✓		✓		
	groups	growed		✓		✓		
	their	there		✓		✓		
	grew	grow	✓		✓			
	alike	alway		✓		✓		
Total		27	6	17	4	19	2	2

C-TEST ERROR CODING SHEET

Name: N·go Form: C1 [✓ MC1] C2 MC2

Subtest	Text	Error	Error Type					
			Substitution				No Response	Unintelligible
			Grammatical Function		Meaning Change			
			Same	Different	Little or No	High		
S1	new	now		✓		✓		
S3	water	waner						✓
	oxygen	forgen						✓
	lungs	–					✓	
	Fortunately	–					✓	
	acquire	–					✓	
	oxygen	forgen						✓
S4	lived	–					✓	
	isolation	–					✓	
	under	other		✓		✓		
	conditions	–					✓	
	centuries	–					✓	
	alike	broke	✓			✓		
Total		13	1	2	0	3	7	3

APPENDIX F

SAMPLE INTERVIEW QUESTIONS

SAMPLE INTERVIEW QUESTIONS

1. Do you like this type of test? (How much?) Why? (Why not?)
2. Do you think it is a good test of your English ability? Why? (Why not?) What do you think it tests?
3. How did you do the test? (Did you read word by word?) When you didn't know the answer, what did you do?

   These options are given if necessary:

   a. Read word by word.
   b. Read the whole sentence first.
   c. Read the whole text first.
4. Do you translate (from English to your language) when you read in English? How much? (How often?)
5. Do you think you are a good reader? Why? (Why not?) How do you read in English? What do you think is most important when you read in English? Why?

   These options are given if necessary:

   a. To pronounce the words correctly.
   b. To know the meaning of the words.
   c. To understand the meaning of the text.